

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

Evaluation Synthesis Report – Vol. 1.2 – Annexes

January 2019

---



Oxford Policy Management



itad

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

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



**UKaid**  
from the British people

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

## Preface

DFID contracted the e-Pact consortium to deliver the Monitoring, Verification and Evaluation (MVE) services for the multi-country WASH Results Programme. The final synthesis report integrates key findings and recommendations from the previous midline evaluation (2016) with the results from the endline evaluation activities conducted from October 2017 to July 2018.

This document contains the annexes to that final synthesis report.

This evaluation is being carried out by e-Pact. Julia Larkin is the team leader for the evaluation and Zach White is manager for the evaluation ([Zach.White@opml.co.uk](mailto:Zach.White@opml.co.uk)). The project director for the whole monitoring, verification, and evaluation contract is Don Brown.

The contact point for the client is Stephen Lindley-Jones ([Stephen-jones@dfid.gov.uk](mailto:Stephen-jones@dfid.gov.uk)). The client reference number for the project is PO6507.

---

e-Pact	Level 3 Clarendon Centre 52 Cornmarket St Oxford OX1 3HJ United Kingdom	Tel +44 (0) 1865 207300 Fax +44 (0) 1865 207301 Email <a href="mailto:admin@opml.co.uk">admin@opml.co.uk</a> Website <a href="http://www.opml.co.uk">www.opml.co.uk</a>
Registered in England: 3122495		

## Table of contents

Preface	1	
Annex A	Executive summary of the midline evaluation report	3
Annex B	Evaluation approach and methodology	7
Annex C	Critique of the theories of change and associated assumptions	49
Annex D	Summary of contribution analysis	57
Annex E	Evidence matrices	64
Annex F	Overview of data collection	69
Annex G	Summary of verification processes	74
Annex H	Stakeholder comments on the report	79
Annex I	VFM analysis based on price	80
Annex K	Original Terms of Reference	97

## Annex A Executive summary of the midline evaluation report

### A.1 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.

### A.2 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.

## **A.3 Key findings**

### **A.3.1 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.

### **A.3.2 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.

## **A.4 Recommendations**

### **A.4.1 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.

#### **A.4.2 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.

#### **A.4.3 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.



## Annex B Evaluation approach and methodology

The midline evaluation report (2016) examined progress half way through the four-year WASH Results Programme and followed the methodology outlined in the Evaluation Design Document. It answered the subset of evaluation questions that 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). It addressed all relevance evaluation questions, and the effectiveness and efficiency questions as they related to the output phase, and it explored some initial indications of impacts and sustainability.

In preparation for the endline evaluation activities, the team developed the Endline Design Note to document updates and refinements to the methodology to be used for the endline evaluation, including: the DEQs to be addressed, our overarching evaluation approach, and our specific methodology for the various components of the evaluation. While the evaluation examines **what** was achieved, the main focus is on **how** and **why**, and **under what circumstances**, results were achieved.

The endline evaluation activities were conducted from October 2017 to July 2018, and addressed the outstanding effectiveness and efficiency elements, especially as they related to the outcome phase (January 2016 to March 2018), and examined the impact and sustainability evaluation questions.

Table 1 summarises the stakeholder engagement and desk review activities conducted at endline, which are almost identical to those of the midline, with two key exceptions. Interviews with unsuccessful suppliers conducted at midline were not repeated, as not only are they not relevant for the endline evaluation questions, but no new information was anticipated from this source. More effort was put into the remote interviews with the country managers for countries not included in the site visits, to increase the robustness and internal validity of results.

**Table 1: Evaluation activities at programme level and country level at endline**

Level of investigation	Evaluation activities
Programme level	Review of supplier documentation and verification reports
	Literature review of PbR
	Interviews with DFID staff
	Interviews with lead suppliers
	Remote interviews with all country managers
	Analysis of verified results data
	Interviews with lead verifiers
	Interviews with L&D team
Country-level investigation in three case studies	Review of country programme documentation
	Interviews with: supplier country staff; other key informants at national level; local implementing partner staff
SAWRP: Pakistan SSH4A: Uganda SWIFT: Kenya	Interviews with: members of targeted communities (service users); other key informants at community level



Refer to Box 1 for an assessment of the extent to which the WASH Results Programme was managed and delivered in line with the Paris Declaration principles.

### Box 1: The extent to which the WASH Results Programme was managed and delivered in line with the Paris Declaration principles<sup>1,2</sup>

Analysis of the WASH Results Programme's alignment with the Paris Declaration principles was not an objective of this evaluation. Thus, such evidence was not sought out and the evaluators did not come across evidence that the suppliers were following these principles explicitly. Nonetheless, the evaluators consider that the WASH Results Programme delivered in line with the Declaration's five principles in the following ways:

**Ownership:** The WASH Results Programme worked within the policy frameworks of developing countries and sought to enhance these by catalysing the WASH in countries.

**Alignment:** As stated above, the WASH Results Programme worked within local systems and national WASH objectives to the extent feasible.

**Harmonisation:** The suppliers of each of the sub-programmes are experienced WASH providers for DFID, and in some cases also had, or were at the time of the WASH Results Programme, implementing WASH programmes for DFID in the same or similar areas. There is potentially room for DFID to further harmonise with other donors to further reduce country-specific barriers to sustainable WASH outcomes.

**Results:** The WASH Results Programme aimed to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people.

**Mutual accountability:** DFID, together with its delivery agents, was accountable for the development results.

## B.1 Summary of changes from midline to endline

To the extent appropriate, the endline evaluation activities followed the approach used in the midline. For example, the same theory of change underpinned the approach, the same evaluation questions were used (with only minor updating), the same data collection activities were conducted, and the three countries selected for site visits in the midline were also visited in the endline.

The key differences or updates are as follows:

### Evaluation questions

- **Modified scope for endline.** Out of the 28 evaluation questions, only 11 were revisited at endline, because in the remainder of cases the evaluation questions were comprehensively addressed during the midline and new data collection was not expected to lead to different findings. However, this report incorporates findings for all evaluation questions, summarising highlights from the midline, as well as the new results from endline evaluation activities, especially as they relate to outcome-level efficiency and effectiveness, as well as impacts and sustainability.
- **Minor rewording.** Some of the evaluation questions were slightly rephrased, in order to update our understanding of DFID's priorities as well as what was feasible to address. The updated wording is captured in the list of evaluation questions in Annex B.2 below.

<sup>1</sup> This analysis was added to bring the Report into full alignment with the requirements of the DFID Evaluation Quality Assurance and Learning Services (EQuALS) Quality Assurance Evaluation Report template, which is often used by DFID and other UK Government departments spending overseas development assistance to assess the quality of evaluations.

<sup>2</sup> For more information on the Paris Declaration and its principles, see the OECD website: [www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm](http://www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm) [Accessed 24 September 2018]

- **Remove/replace two evaluation questions.** There were also two evaluation questions under 'efficiency' where our evaluability assessment showed that they are no longer appropriate as worded due to data availability (indicated as 'removed' in the table below).
- **Addition of one evaluation question.** We added a new evaluation question relating to risk transfer.

## Approach

- **Increased focus on data collection for countries where no country visit took place.** Instead of carrying out community-level impact assessment in case study countries, the evaluation team and DFID agreed that these resources could be more effectively used (i.e. added value for answering multiple evaluation questions) for additional interviews with the countries where a country visit was not originally planned. This approach was tailored as appropriate to the context and opportunities of each country targeted.
- **Shift away from VFM.** In consultation with DFID, it was confirmed that only qualitative aspects of efficiency would be assessed, due to challenges in obtaining secondary data on programme spending from suppliers to undertake a full VFM analysis.
- **Analytical frameworks for impacts, sustainability, and inclusivity adapted.** The risk-based analytical framework outlined in the original Evaluation Design Document was adapted to spell out in more detail the 23 prerequisites for achieving sustainability, eight prerequisites for achieving inclusivity, and four key prerequisites for achieving health impacts.
- **Contribution analysis.** The contribution analysis originally planned for the midline stage was postponed to this endline stage when the outcome phase was complete and prospects for impacts would be more apparent. Refer to Annex E below for more details.

## Team composition

The evaluation team comprises experienced evaluators who regularly work with DFID and there is no known conflict of interest.

The majority of core staff were the same from midline to endline. The endline core team included Lucrezia Tincani, Zach White, and Jeremy Colin, who had also been involved in the midline, with the addition of Sue Cavill. The Team Leader role shifted to Julia Larkin. Two senior members from the midline (Richard Carter and Timothy Forster) shifted to advisory positions for the endline. One of the benefits of this shift was that it allowed several country visits to be carried out by the same WASH experts, thus minimising the risk of silos between suppliers.

The core team conducted the data collection activities, analysis, and report drafting. The advisory team was engaged at key intervals throughout the evaluation to: help shape the approach; review the analysis at a few critical break points (e.g. following the DFID/supplier/verifier roundtable consultations); provide sense checks, as well as advice during the evaluation; vet the emerging lessons and recommendations; and review all deliverables. Given their prior engagement with the midline, they also helped ensure continuity.

## B.2 Evaluation questions

Table 2 lists the evaluation questions that were addressed at the midline and/or endline stage. As noted above, one evaluation question on risk transfer was added, two efficiency evaluation

questions were removed, and the wording was updated for five evaluation questions. For those questions for which the wording was updated for the endline, the new wording is presented here.

**Table 2: Evaluation questions**

Evaluation question	Midline	Endline	Evaluation method
<b>HEQ1 – relevance:</b> <i>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?</i>			
DEQ 1.1: To what extent were the <b>programme objectives</b> clearly articulated?	✓		Document review
DEQ 1.2: To what extent does the <b>programme's design</b> (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?	✓		Document review
DEQ 1.3: To what extent were the <b>scale and pace of the programme</b> (including the December 2015 deadline) realistic for achieving intended outputs and outcomes given the capacity of the suppliers and their local partners?	✓		Document review
DEQ 1.4: To what extent was <b>the PbR modality</b> appropriate for achieving sustainable and inclusive WASH outcomes, given the capacity of the suppliers and the timeline of the programme?	✓		Process evaluation
DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?	✓		Critique of TOC
DEQ 1.6: How likely was it that the programme design would encourage suppliers to propose 'innovative WASH interventions'?	✓		Critique of theory of change
DEQ 1.7: How likely was it that the programme design would encourage inclusive outputs and outcomes?	✓		Critique of theory of change
DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?	✓		Critique of theory of change
DEQ 1.9: To what extent was the <b>design of each consortium sub-programme</b> appropriate for achieving DFID's key objectives?	✓		Critique of theory of change
<b>HEQ2 – effectiveness:</b> <i>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?</i>			
DEQ 2.1: Did the programme achieve the intended outputs at scale?	✓		Process evaluation
DEQ 2.2: To what extent have the prerequisites for achieving inclusive WASH outcomes been addressed by suppliers? <sup>3</sup>		✓ <i>Wording updated</i>	Contribution analysis
DEQ 2.3: To what extent have services continued to function and have behaviours continued to be used since their initial implementation (sustainable outcomes)?		✓	Process evaluation
DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	✓	✓	Contribution analysis
DEQ 2.5: Under which circumstances did the <b>PbR framework</b> help/hinder the achievement of intended outputs and outcomes?	✓	✓	Contribution analysis
DEQ 2.6: Under which circumstances did the <b>PbR framework</b> affect the quality of programme implementation (positive or negative)?	✓	✓	Contribution analysis

<sup>3</sup> Inclusive WASH outcomes are defined as not excluding any section of society or target area.

DEQ 2.7: Under which circumstances did suppliers implement innovative approaches and focus on learning?	✓	✓	Process evaluation
<b>HEQ3 – efficiency: Has the programme been designed and implemented in a cost-efficient manner?</b>			
DEQ 3.1: How efficient was the <b>tendering and procurement process</b> and what effect did this have on programme delivery?	✓		Process evaluation
DEQ 3.2: To what extent were the <b>individual sub-programmes</b> designed and delivered in a cost-efficient and cost-effective manner?		Removed	n/a
DEQ 3.3: Under which circumstances did the <b>PbR modality</b> affect the cost-efficiency and cost-effectiveness of individual sub-programmes?	✓	Removed	n/a
DEQ 3.4: To what extent did the <b>PbR modality</b> strengthen the programme monitoring of individual sub-programmes?	✓	✓ Wording updated	Contribution analysis
DEQ 3.5: Under which circumstances did <b>key programme features</b> affect cost-efficiency and cost-effectiveness?	✓		
DEQ 3.6: How did the efficiency of the <b>programme management arrangements</b> of individual sub-programmes affect programme delivery?	✓	✓ Wording updated	Process evaluation
DEQ 3.7: To the extent new PbR <b>risk-sharing</b> arrangements were applied within consortia, how did this affect programme delivery?		✓ (New EQ)	Process evaluation
<b>HEQ4 – impact: How likely is it that the programme will achieve its final impact objectives while minimising unintended negative consequences?</b>			
DEQ 4.1: How likely is it that the programme will achieve its health and non-health <b>impacts</b> ?		✓	Process evaluation
DEQ 4.2: Under which circumstances did the WASH Results Programme activities have any <b>unintended/ unplanned positive or negative impacts</b> ?		✓	Impact assessment
<b>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?</b>			
DEQ 5.1: To what extent were the <b>individual sub-programmes</b> designed and implemented to maximise the likelihood of achieving long-term sustainable WASH outcomes and impacts?	✓	✓	Contribution analysis
DEQ 5.2: Under which circumstances has the <b>PbR modality</b> affected the likelihood of long-term sustainability of the outcomes and impacts?		✓	Contribution analysis
DEQ 5.3: Under which circumstances have <b>other programme features</b> affected the likelihood of the long-term sustainability of the outcomes and impacts?		✓	Process evaluation
DEQ 5.4: Under which circumstances did the WASH Results Programme contribute to enhanced sector learning to inform <b>better evidence-based WASH policy and programming</b> ?		✓ Wording updated	Process evaluation

## B.3 General approach by evaluation question group

### Assessing relevance

The relevance of the WASH Results Programme was comprehensively assessed at midline, and was not proactively revisited at endline.

At midline, assessing the relevance of the WASH Results Programme was structured around a document review, as well as a critical analysis of the assumptions underlying the programme-level theory of change, to determine the strengths and weaknesses of the programme design. Analysis first involved assessing whether the programme's strategic objectives were valid and appropriate

to meet supplier and sector needs. Second, it was assessed whether the activities planned were appropriate for achieving these objectives. The data collected, as described above, especially the programme documentation and interviews with lead suppliers, were analysed to address the remaining relevance DEQs.

At endline, we conducted another high-level review to ensure the programme-level theory of change, as well as the two nested theories of change, were still accurate. In a few minor cases the evaluation team gained additional context from the data collected during the endline relating to the outcome phase, which is incorporated as noted in Annex G.

## Assessing effectiveness

The evaluation used a theory-based approach to explore the programme's effectiveness. This approach entailed using the theory of change of the programme as a starting point to test the assumptions implicit in DFID's original programme design.<sup>4</sup> In other words, we explored: (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 results, and the reasons why. To the extent feasible, we sought information on other WASH programmes being implemented by the same suppliers. The verification reports provided insight into *whether* planned outputs and outcomes have been achieved – therefore the evaluation focused on *how and why* the programme has (or has not) achieved its intended results.

The endline evaluation activities gathered detailed information on the causal factors and contextual factors affecting outcome achievement (the midline focused on outputs only). Beyond review of programme documentation, these factors were assessed primarily through two methodologies:

1. **Process evaluation**, including supplier interviews and country visits, was used to assess how implementation activities had been implemented and in which ways these contributed to programme effectiveness.
2. **Contribution analysis** was used for a small number of evaluation aspects, to draw causal inferences to explain how change was generated, based on a plausible theory of change and recognising that the programme was unlikely to be the sole factor producing change but that it did so in interaction with contextual factors. Specifically, contribution analysis was used for DEQs 2.2, 2.4, 2.5, 2.6, 3.4, 5.1, and 5.2, because it helped tease out the factors that contributed to the quality of implementation and achievement of intended outputs and outcomes. Annex F provides more detail on the contribution analysis approach.

## Conceptual grounding for inclusion assessments

Governments, donors, international agencies, and NGOs are concerned that the benefits of WASH services should reach everyone, irrespective of gender, age, disability, health, ethnicity, religion, or social status. In WaterAid's Equity and Inclusion Framework (Gosling, 2010) a distinction is made between **equity** – '*the principle of fairness*' – and **inclusion** – '*the process for ensuring that all are able to participate fully*'. The pursuit of equity requires a recognition of the different needs of different people, and actions to compensate for discrimination and disadvantage. The process of inclusion involves '*... supporting people ... to engage in wider processes to ensure that their rights and needs are recognised*'. According to Gosling, achieving equity and inclusion requires:

<sup>4</sup> The evaluation team based its evaluation approach on two theories of change: DFID's programme-wide theory of change and a second theory of change constructed by the evaluation team to provide further detail on the underlying PbR modality.



- *better recognition and understanding of the differential needs of individuals and groups;*
- *identifying and tackling the root causes of exclusion;*
- *promoting and supporting their inclusion in decision-making processes; and*
- *identifying and implementing appropriate and sustainable solutions.*

Addressing equality and inclusion requires action at multiple levels. Projects and programmes have target populations, all of whose WASH needs must be identified and addressed. Discriminatory attitudes towards those who are commonly excluded also need to be addressed, widening the circle of actions beyond the target population. Government policies, guidelines, and budget allocations also need to reflect a positive attitude to meeting diverse needs.

In this vein, Gosling (2010) highlights three barriers to inclusion which must be overcome:

- **negative attitudes**, such as prejudice, pity, isolation, overprotection, stigmatisation, misinformation, and family shame;
- **‘environmental’ barriers**, including difficulties of physical accessibility and those aspects which present particular difficulties for specific groups (e.g. access to visual communication by those with visual impairments);
- **institutional / organisational barriers**, including policies and processes which exclude or neglect those with particular needs, whether deliberately or not.

All three of these barriers have to be addressed in order to achieve truly equal and inclusive outcomes. The first and third are arguably the most difficult and time-consuming to change since they involve cultural and institutional change. Wilbur and Jones (2014) provide some guidance on this matter, however. The second can at least be addressed in the short term through programmes which aim to reach the entirety of their target populations. For example, recent guidance on WASH technologies which are accessible to disabled and older people has been published by WEDC, WaterAid, and Share (Jones and Wilbur, 2014).

Annex D.2 of the Evaluation Design Document discusses these conditions with regard to the programme’s logframe and provides a discussion of how these principles can be translated into action in programming. The results of this work at the design stage was a set of 12 actions across the areas of: programme design and implementation; addressing institutional barriers; and monitoring. Ahead of the endline assessment, and formalised through the Endline Design Note, these were further revised into an assessment framework focusing on eight key questions against which WASH Results Programme implementation was considered.

The team discussed with the suppliers sources that could inform assessments of inclusion, and it was determined that assessing *whether* the WASH Results Programme has delivered inclusive results cannot be directly measured through secondary programme data. Whilst SNV does collect some data, the other two suppliers did not report their findings disaggregated by population wealth quintiles and gender. Given these limitations, we referred to the verification reports, which give some indication of inclusion risks.

## Inclusion framework

A **process evaluation** approach was used to assess to what extent the WASH Results Programme addressed the prerequisites for inclusion (see DEQ 2.2). This evaluative judgement was developed based on the presence or absence of key elements outlined in the **risk-based analytical framework**. Note: to aid the flow of the report, the inclusion discussion is included under impacts in the main body of the report (Section 4.4).

The framework ultimately developed for assessing *which factors* affected inclusion, and whether the WASH Results Programme sufficiently addressed the prerequisites for equality, is presented in Section 4.4.2 of the main report. Evidence primarily come from:

- mid-term review findings (especially on how design and implementation addressed equity and inclusion);
- verified outcome data and findings of other relevant surveys commissioned during the outcome phase;
- programme documentation, including suppliers' outcome-phase strategies/plans and progress reports;
- findings of country visits (case study countries); and
- interviews with programme managers and verifiers (countries not visited).

For the case study countries, interviews were used to compare the perspectives of different stakeholders, including supplier staff, government partners, lead and country verifiers, and neutral informants (such as other development agencies active in the sector). This facilitated our ability to triangulate across different perspectives and minimise any biases.

**Table 3: Framework for assessing whether the prerequisites for inclusive WASH outcomes have been addressed by suppliers**

Areas of investigation	Achievement		
	Unlikely	Somewhat	Yes/likely
<i>[Numbers in this row are scores based on professional judgement]</i>			
<b>Programme planning and implementation</b>			
1. Within targeted locations, did the programme endeavour to meet the needs of all, including communities that were harder to reach or serve?			
2. Was technology selection (where relevant) and detailed design undertaken with the full participation of the intended beneficiaries?			
3. Within targeted communities, did operational approaches address the needs of marginalised groups/households and of those with physical disabilities and infirmities?			
4. Did women participate actively in programme implementation and were they adequately represented in decision-making processes?			
<b>Monitoring</b>			
5. Did monitoring at output level generate disaggregated beneficiary data confirming that the programme provided access to WASH facilities for marginalised groups and those with special needs?			
6. Did outcome-phase surveys confirm the use of WASH facilities and adoption of hygienic behaviour by marginalised groups and those with special needs?			

## Assessing efficiency

As was discussed with DFID, it was challenging for the evaluation to assess *whether* efficiency has been achieved through VFM analysis due to the sensitivities involved in obtaining secondary data on programme spending from suppliers.<sup>5</sup> Therefore, DEQs 3.2 and 3.3 were removed.

<sup>5</sup> C.f. email exchange with Guy Howard on 16 December 2016.



The VFM analysis carried out at midline on price data (not cost data) was repeated to compare price per beneficiary (for outcomes) as budgeted at programme start, with the price per beneficiary achieved (for outcomes) by programme end. This analysis was not included in this endline evaluation report but was submitted to DFID in a separate report for confidentially reasons.

Overall, the endline evaluation focused on *qualitative* aspects of efficiency, namely the effect of the PbR modality on programme monitoring (DEQ 3.4) and the implications of programme management arrangements (DEQ 3.6). We added a new evaluation question exploring how variations in risk-sharing across the three suppliers affected programme delivery (new DEQ 3.7).

## Assessing the likelihood of impacts

### Conceptual grounding of the approach to evaluating the prospects for health impacts

There is currently a strong emphasis among those funding and implementing development programmes on the achievement and measurement of beneficial development impacts. In the case of WASH, some of the most important impacts are thought to be on health, especially (but not only) the reduction of under-five diarrhoeal morbidity and mortality. The discussion here is written in the context of (a) the implementation of WASH programmes by international NGOs working in partnership with local NGOs and local and national governments, and (b) the need to evaluate the sustainability of the interventions undertaken.

DFID's Evidence Paper (DFID, 2013) provides a succinct summary of the health impacts of WASH interventions and their possible magnitude. The analysis in that paper is rooted in two key conceptual frameworks: i) Bradley's classification of water-related illnesses, modified by Dar and Khan (2011); and the 'F' diagram, first proposed by Wagner and Lanoix, 1959. The Bradley classification is helpful in understanding the ways in which different aspects of WASH interventions are more or less relevant to different categories of disease. It is worth noting, however, that the DFID evidence paper highlights the impacts of combined improvements in water supply, sanitation, and hygiene do not seem to be additive – the authors concluded that this is because all interventions to a greater or lesser extent combine these different aspects. The F diagram focuses specifically on those diseases having faecal–oral transmission routes, and is a useful frame for understanding the ways in which WASH and other interventions (e.g. food hygiene – not explicitly highlighted in the figure) can break transmission pathways<sup>6</sup>.

The achievement of WASH health impacts depends first on the delivery of programme outputs, and then on their continued functioning and utilisation (programme outcomes). Furthermore, since WASH programmes attempt to (a) remove human faeces from the environment, (b) promote better hygiene, and (c) ensure that there is sufficient water to permit effective sanitation and hygiene, it is important that WASH results are enjoyed by all – open-air defecation or lack of handwashing by

<sup>6</sup> Key conclusions of the Evidence Paper with regards to health impacts include the following: based on suggestive evidence, 7% of the total disease burden could be prevented annually with safe WASH; there is good evidence that the health impacts of WASH extend far beyond diarrhoea to include many other important diseases, including acute respiratory infections, undernutrition, soil-transmitted intestinal helminth infection (ascaris, trichuris and hookworm), schistosomiasis, Guinea worm, trachoma and certain non-infectious diseases associated with chemical water quality (arsenicosis and fluorosis); there is good evidence that improved hygiene can reduce the risk of acute respiratory infections (pooled estimate of 23% risk reduction); there is suggestive evidence that diarrhoea (as a consequence of inadequate WASH) contributes to undernutrition, but there is currently weak evidence for causal links to tropical enteropathy; there is an unequal distribution of WASH-associated mortality and morbidity with disproportionately high levels among the poorest populations and the majority of deaths from diarrhoea occur among children.

one individual may have a negative health impact on another. WASH services and behaviours are as much public as private goods.

The first prerequisites for the achievement of health impacts are therefore:

1. the sustained functioning and utilisation of those services and behaviours (i.e. with no time limit); and
2. the utilisation of water and sanitation services and hygiene practices by all members of target populations.

Two further things need to be in place to see significantly reduced levels of water- and sanitation-related disease. (The fact that they are absent in low-income and fragile states is a partial explanation for the limited health impacts of WASH interventions on their own.) First, all members of the entire population need to utilise improved WASH services and practise good hygiene. If they do not, then disease which is prevalent in less well-served communities will spill over into those which are served. Second, other complementary measures need to be in place, including those focused on food security and nutrition, shelter, income, access to immunisations and health care in general, and access to education (especially by girls and women. The third and fourth prerequisites for achieving health impacts in the context of this programme are therefore seen to be:

3. that WASH interventions at limited scale nevertheless advocate for accelerating national WASH programmes, including advocating for adequate and appropriately targeted financial resources; and
4. that communities experiencing better WASH services also enjoy adequate food security and nutrition, shelter, income, access to immunisations, and healthcare in general, and access to education (especially by girls and women).

In the context of this evaluation these prerequisites are framed around two 'backward-looking' outcome-level questions (corresponding to prerequisites 1 and 2 above):

*Q1: To what extent have services and behaviours continued to function and be used since their initial implementation (sustainable outcomes)?*

*Q2: 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)?*

Recognising that the full achievement of both outcomes and health impacts may post-date the programme duration, and that an evaluation may have to be carried out before the full realisation of those results, there is value too in asking 'forward-looking' questions:

*Q3: To what extent has the programme advocated for, and successfully influenced, attempts to bring sustainable WASH services to the entire population?*

*Q4: What obstacles exist to the realisation of the full potential health benefits of the WASH programme, in sectors such as food, shelter, livelihoods, health, and education?*

At the midline there was further discussion between e-Pact and DFID regarding the framing of health impacts. In particular, the evaluation team was asked to further justify two propositions, namely our contentions that limited health impacts are to be expected (a) if water, sanitation and hygiene are not implemented in an integrated manner, and (b) if sanitation coverage (and use) do not reach some (high) threshold level in a given community. An abridged version of the evaluation team's response is summarised below:

- Regarding the additive effects of joint implementation of water, sanitation, and hygiene programme components:** It was acknowledged that the systematic reviews of health impact from various aspects of WASH have all failed to show an additive or cumulative effect of joint implementation of water, sanitation, and hygiene programme components. To explain this the following view is expressed in DFID's 2013 Evidence Paper<sup>7</sup>: '*... a number of authors have commented that the health impact of combined interventions such as sanitation plus water supply is no greater than the impact of a single intervention. This apparent anomaly becomes clear when it is understood that the comparison is not between studies of combined interventions and studies of single interventions, but between studies of combined interventions and other studies of combined interventions.*'<sup>8</sup> In other words we may not be able to distinguish clearly between the individual contributions of water, sanitation, and hygiene because all interventions considered in the few studies which contribute to the systematic reviews included some elements of each. This view reflects earlier work<sup>9</sup> that sought to address the same question, which went further to also suggest '*[the] logic and understanding of the F-Diagram and the major transmission routes for faecal-oral diseases would suggest that each should have an independent effect on the transmission and prevalence of this disease classification, including diarrhoeal infections*'.

Brown *et al.* 2011<sup>10</sup> comment that 'Recent research [citing Hunter *et al.* 2009<sup>11</sup>] has shown, however, that even such improved water supplies may be subject to faecal contamination (ibid.) and that even occasional exposure to unsafe water—for example, from intermittent service or inadequate treatment—can undermine health benefits.' The reference here is to breakdowns in water services. However, the same argument applies to consumption of contaminated drinking water even where sanitation access and usage are high.

- Regarding the threshold level of sanitation usage in communities:** Only recently has there been a growing body of literature using experimental methods<sup>12</sup> that has examined the link between rural sanitation programmes and health. Several key studies (Clasen *et al.* 2014;<sup>13</sup>, Patil *et al.* 2014;<sup>14</sup> and Pickering *et al.* 2015<sup>15</sup>) failed to identify health impacts. Given the strong biological plausibility of health impacts a key explanation put forward by the authors for the null results in the trials was the notion of a 'threshold coverage' as the coverage levels achieved in the interventions studied were below full coverage, though it was also acknowledged by some authors that 'there is inadequate evidence from previous research to support this hypothesis.' Duflo *et al.* (nd)<sup>16</sup> undertook a study of the health impact of a programme which provided universal access within a village to hygienic latrines and in-home piped water at the same time, at a reasonable cost. They estimated reduction in episodes of severe diarrhoea of 30–50%, compared to a null effect on diarrhoea shown in the then only two RCTs on sanitation (Pickering's study followed, and showed the same finding in relation to sanitation only). The

<sup>7</sup> DFID (2013) 'Water, Sanitation and Hygiene Evidence Paper, May 2013.

<sup>8</sup> Page 45.

<sup>9</sup> WELL FACTSHEET Health Impacts of Improved Household Sanitation. Author: Beth Scott, Quality Assurance: Sandy Cairncross and Andrew Cotton. November 2006

<sup>10</sup> Brown J, Cairncross S, Ensink JHJ (2011) *Arch Dis Child* doi:10.1136/archdischild-2011-301528

<sup>11</sup> Hunter PR, Pond K, Jagals P, *et al.* (2009) 'An assessment of the costs and benefits of interventions aimed at improving rural community water supplies in developed countries'. *Sci Total Environ* 2009;407:3681–5.

<sup>12</sup> Predominantly cluster RCTs.

<sup>13</sup> Clasen T *et al.* (2014) 'Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial'. *Lancet Glob Health*. 2014 Nov; 2(11):e645-53. doi: 10.1016/S2214-109X(14)70307-9. Epub 2014 Oct 9.

<sup>14</sup> Patil SR, *et al.* (2014) 'The Effect of India's Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial'

<sup>15</sup> Amy J Pickering, Habiba Djebbari, Carolina Lopez, Massa Coulibaly, Maria Laura Alzua (2015) 'Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial', *Lancet Glob Health*. 3: e701–11

<sup>16</sup> Duflo E, Greenstone M, Guiteras R, Clasen T (nd) 'Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation'

authors comment as follows: ‘These positive and large results differ from the null findings of two recent cluster-randomised trials evaluating sanitation-only interventions in rural India, one in Orissa (Clasen *et al.* 2014) and the other in Madhya Pradesh (Patil *et al.* 2014). We do not view our results as contradictory, for two reasons. First, RHEP is a package of a communal water tank, piped water to the home, and household latrines and bathing facilities. This evaluation therefore does not provide information on the effectiveness of the individual components, but the contrast with the RCT results suggests that this ‘full package’ is important. Second, latrine use was relatively low in the two RCTs, which likely limited their health impact (Brown and Clasen 2012). We speculate that providing a tap for piped water at the toilet reduced barriers to maintenance and hygienic use, but further research is needed to understand the factors that motivate or impede use of sanitation facilities (Coffey *et al.* 2014, Gertler *et al.* 2015).’

- **Summary.** The absence of water supply improvements, which are necessary for the practise of safe sanitation and good hygiene interventions, would reasonably be expected to limit health impacts. Undertaking sanitation and hygiene interventions only, in communities which lack adequate water supply (as seen in this evaluation), represents at the very least a lost opportunity to achieve a significant health impact. Our contention that the diarrhoeal disease impact of Lot B interventions is likely to be insignificant is borne out by three recent RCTs and the study by Duflo *et al.*, among others.

Likewise, levels of latrine usage in a community that are not sufficiently high could also fail to impact sufficiently on faecal contamination of the environment, so limiting the potential health effects of the programme. Although there is insufficient rigorous evidence to categorically conclude that a certain general threshold level of sanitation usage is necessary for health impact, all modern sanitation programming designs for total sanitation and ODF status; and in cases where health impacts are absent, researchers often explain this by reference to lower than 100% levels of access or latrine usage.

To quote Clasen *et al.* (2014): ‘As efforts to expand sanitation coverage are undertaken worldwide, approaches need to not only meet coverage-driven targets, but also achieve levels of uptake that could reduce levels of exposure, thereby offering the potential for genuine and enduring health gains.’

The proposition that there is a threshold level of coverage has prompted further research seeking to address that particular question, though no systematic summaries of the evidence exist in this area. Recent interest in the topic has spurred further research in the area and several studies published since the midline evaluation report add to the discussion. Okullo *et al.* (2017)<sup>17</sup> conducted a study where they compared water quality between ODF-certified and non-ODF-certified villages in Kenya and found that ODF status has a significant impact on water quality. Harris *et al.* (2018) conducted a study in Mali using ‘index households’ to assess the relationship between the household latrine ownership and latrine ownership of the surrounding households; they concluded that child height-for-age had a significant and positive linear relationship with community latrine coverage, while child weight-for-age and household water quality had nonlinear relationships. Child growth and water quality were not associated with individual household latrine ownership. That is, the latrine status of the surrounding households was a more significant factor in child health than the latrine ownership status of the household. Aiming for community-wide impacts has significance with regards to the mechanisms of behaviour change (collective social pressure)

<sup>17</sup> See Okullo, J. O., Moturi, W. N., & Ogendi, G. M. (2017). Open Defaecation and Its Effects on the Bacteriological Quality of Drinking Water Sources in Isiolo County, Kenya. *Environmental Health Insights*, 11, 1178630217735539. <http://doi.org/10.1177/1178630217735539>



and equity considerations<sup>18</sup> (focusing on community-wide impacts by definition is more inclusive), and increasingly there is evidence that it is significant to ensuring health impacts.

Regarding the question of the additive effects of joint implementation of water, sanitation, hygiene, and nutrition programme components the results of the WASH Benefits trial provide an addition to the evidence. The WASH Benefits trial was conducted in Kenya and Bangladesh and enrolled pregnant women and conducted one- and two-year follow-up assessments with women allocated to arms that received single or multiple WASH and nutrition interventions. In rural Kenya (Null *et al.* 2018<sup>19</sup>) the study concluded *'none of the interventions reduced diarrhoea prevalence compared with the active control. Compared with active control (length-for-age Z score -1.54) children in nutrition and combined water, sanitation, handwashing, and nutrition were taller by year 2 (mean difference 0.13 [95% CI 0.01–0.25] in the nutrition group; 0.16 [0.05–0.27] in the combined water, sanitation, handwashing, and nutrition group). The individual water, sanitation, and handwashing groups, and combined water, sanitation, and handwashing group had no effect on linear growth'*. In Bangladesh (Luby *et al.* 2018<sup>20</sup>) the study concluded: *'Compared with a prevalence [...] in the control group, 7-day diarrhoea prevalence was lower among index children and children under 3 years at enrolment who received sanitation (61 [3.5%] of 1760; prevalence ratio 0.61, 95% CI 0.46–0.81), handwashing (62 [3.5%] of 1795; 0.60, 0.45–0.80), combined water, sanitation, and handwashing (74 [3.9%] of 1902; 0.69, 0.53–0.90), nutrition (62 [3.5%] of 1766; 0.64, 0.49–0.85), and combined water, sanitation, handwashing, and nutrition (66 [3.5%] of 1861; 0.62, 0.47–0.81); diarrhoea prevalence was not significantly lower in children receiving water treatment (90 [4.9%] of 1824; 0.89, 0.70–1.13). Compared with control (mean length-for-age Z score -1.79), children were taller by year 2 in the nutrition group (mean difference 0.25 [95% CI 0.15–0.36]) and in the combined water, sanitation, handwashing, and nutrition group (0.13 [0.02–0.24]). The individual water, sanitation, and handwashing groups, and combined water, sanitation, and handwashing group had no effect on linear growth'*. The findings are consistent across both countries that nutrition counselling and supplementation had a positive effect on growth (in Kenya combined with WASH), though neither study identified additive benefits from the WASH programme component integration. As with previous research in the area these null results have sparked some debate in the area, with some authors (Coffey and Spears 2018<sup>21</sup>) pointing to observational data that raise questions over the effect sizes the studies were powered to detect and questions surrounding neighbourhood coverage influences; and others (Cummings and Curtis 2018<sup>22</sup>) raising questions surrounding external validity and the specific choice of interventions; while at the same time acknowledging that these results are powerful evidence and highlight key knowledge gaps regarding our understanding of how interventions prevent transmission. In the authors' reply to these letters (Arnold *et al.* 2018<sup>23</sup>) the issues of external validity and context were acknowledged alongside the propositions that the effectiveness of the interventions may vary across contexts in line with baseline access levels and the degree to which the interventions achieve community-wide impacts. The authors finish by highlighting: *'we feel strongly that these findings should not diminish*

<sup>18</sup> See especially Cronin *et al.* (2017) for a discussion of the importance of focusing on community-wide impacts in reference to the SDGs – authored by senior global WASH leaders within UNICEF. <https://equityhealth.biomedcentral.com/track/pdf/10.1186/s12939-017-0709-5>

<sup>19</sup> See Null *et al.* (2018). 'Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial' *Lancet Global Health*; Volume 6, Issue 3, PE316–E329. [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(18\)30005-6/fulltext?code=lancet-site](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30005-6/fulltext?code=lancet-site)

<sup>20</sup> See Luby *et al.* (2018). 'Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial' *Lancet Global Health*; Volume 6, Issue 3, PE302–E315. [https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X\(17\)30490-4.pdf](https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X(17)30490-4.pdf)

<sup>21</sup> Coffey and Spears (2018). [https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X\(18\)30225-0.pdf](https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X(18)30225-0.pdf)

<sup>22</sup> Cummings and Curtis (2018). 'Implications of WASH Benefits trials for water and sanitation'; Volume 6, e613–e614; [https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X\(18\)30192-X.pdf](https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X(18)30192-X.pdf)

<sup>23</sup> Arnold *et al.* (2018). 'Implications of WASH Benefits trials for water and sanitation – Authors reply'. Volume 6, e616–e617; [https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X\(18\)30229-8.pdf](https://www.thelancet.com/pdfs/journals/langlo/PIIS2214-109X(18)30229-8.pdf)

*ongoing, ambitious efforts to achieve the UN Sustainable Development Goals (SDGs): myriad health, equity, and ethical arguments motivate elimination of open defecation and ample supply of microbiologically safe water, even in the absence of a strong link to child growth’.*

## Health framework

A theory-based approach was used to explore the programme’s likelihood of impact, which the programme logframe defines as a 10% reduction against baseline for under-five mortality and the proportion of the population affected by diarrhoeal disease. This approach entailed exploring (i) whether the programme could reasonably be expected to deliver its intended impacts given the way it was designed and structured, and (ii) the extent to which it was likely to do so, and the reasons why.

Our null hypothesis was that the WASH Results Programme would achieve similar health impacts to other WASH programmes of similar scope and scale. To explore this hypothesis, a **process evaluation** approach was used to assess the contextual factors which would affect the likelihood of impacts.

Many factors contribute to the presence or absence of water-related disease and most WASH programmes only affect some of them, and/or operate on too small a scale to remove risks to health across entire districts or regions. (People living in communities where WASH has improved may still be at risk from neighbouring communities where conditions have not changed).

With this in mind, the assessment of the prospects for health impact focused on four questions:

1. To what extent have services and behaviours continued to function and be used since their initial implementation (sustainable outcomes)?
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)?
3. To what extent has the programme advocated for, and successfully influenced, attempts to bring sustainable WASH services across entire districts (or beyond)?
4. What other obstacles exist to the realisation of the full potential health benefits of the WASH programme, in areas such as nutrition, shelter, livelihoods, and education?

Assessments on sustainability and equity/inclusion were a reference point, as were verified outcome data and the findings of other surveys and studies carried out during the outcome phase. However, the evaluation was only be able to assess the prospects in fairly broad terms, and to highlight any critical risks and challenges identified.

## Assessing sustainability

### Conceptual grounding of the approach to evaluating the prerequisites for sustainable WASH outcomes

The approach taken to evaluating sustainability was rooted in the context of (a) the implementation of WASH programmes by international NGOs working in partnership with local NGOs and local and national governments, and (b) the need to evaluate the sustainability of the interventions undertaken. However, most of the content is also applicable to the situation of implementation by local governments, either directly or through outsourcing. It is also relevant to those without an immediate need to undertake evaluation activities.

A sustainable service most fundamentally is interpreted as behaviour or practice which persists over time. However, to place a time limitation on sustainability – ‘*the service was / will be sustainable for 10 years*’, for example – is unhelpful. Sustainability is certainly about the time dimension of services and practices, but no limit must be placed on that time dimension. Once people enjoy a better water service, better sanitation, or an improved set of hygiene behaviours there must be no going back. If the quality of the service or practice improves over time, this is to be applauded; slippage to an inferior level, quality, or usage of service is not. Studies and monitoring of service performance, functionality, usage, and behaviours by their nature focus on the past and the present. The indicators which a regulator uses to examine utility performance reflect the quality of services delivered over the previous year. Reporting of the functionality of rural water-points gives a snapshot in time. Even longitudinal studies can only examine the past and present, albeit over a time period. However, if we are to fully address the matter of sustainability, we need to ask two key interrelated questions:

- What is the **likelihood** that services or practices will persist for the indefinite future?
- What are the **risks and obstacles** to sustainable services or practices, and how can they be mitigated?

A distinction is therefore made between indications that services or practices have been **sustained** to the present date, and the way in which we might determine whether services and practices are indeed **sustainable** over the foreseeable future. To assess or evaluate sustainability we need to **look back** at past performance and **look forward** in the way just alluded to. The indicators used in looking-back analyses still need refinement and improvement, especially in relation to rural (as opposed to urban) water and sanitation services, and generally in relation to hygiene practices. In regard to forward-looking analyses of sustainability, very little work has been done on suitable methodologies for these. Both aspects are addressed in this note, but more emphasis is placed on forward-looking than backward-looking studies.

**Looking back.** A long tradition in rural water and sanitation evaluation (first set out by the World Health Organization in their Minimum Evaluation Procedure in 1983) has focused on two aspects of water and sanitation services: **functioning** (does it work?) and **utilisation** (is it used?). These two high-level indicators of **outcome** have stood the test of time. ‘Looking back’ studies can evaluate the extent to which services and behaviours are functional and utilised. However, it is through understanding the conditions which need to be in place to achieve sustainable change (i.e. the measures that are needed to bridge the gap between outputs and outcomes) that forward-looking sustainability assessments can be undertaken. Annex D.2 of the Evaluation Design Document discusses these facets with reference to the WASH Results Programme theory of change.

**Looking forward.** Of the various WASH sector initiatives and publications related to sustainability<sup>24</sup> only a small number coherently set out the prerequisites for sustainable WASH outcomes. In addition, many of the ‘tools’ reviewed by Schweitzer *et al.* (2014) contain implicit sets of prerequisites, mostly in the form of indicators of multiple dimensions of sustainability. Schweitzer *et al.* (2014) identified around 220 ‘tools’ which have been developed to assess or predict sustainability. They then undertook a more detailed analysis of 25 of them. Typically, these tools contain ‘areas’ or components (the usual set of social, technical, financial, institutional, and environmental or similar), each of which is characterised by a number of indicators and sub-

<sup>24</sup> See addendum to Annex D.2 of the design document for a mapping of initiatives and key publications related to WASH sustainability.



indicators to produce composite scores. The tools can be applied at project or programme level, or at sector level.

The framework developed for this evaluation is an attempt to synthesise the prerequisites into a single list, combining elements that consider backward-looking and forward-looking elements. The framework also distinguishes between implementation/institutional levels, the fundamental logic being that prerequisites at the user / community level all need to be reinforced at the level of the local administration (together with the local private sector and/or NGOs). Similarly, all actions undertaken at local government level must be supported by the actions of higher-level government organisations and foreign/international NGOs, agencies, and donors. In the framework, repetition is avoided by listing only the highest priority actions at levels higher than the local community.

## Framework for sustainability assessments

A **process evaluation** approach was used to assess the contextual factors that affect the likelihood of the programme results being sustainable beyond programme end. The evaluative judgements regarding the sustainability and inclusion of the WASH Results Programme were based on the presence or absence of key elements outlined in the risk-based analytical framework. In short, the evaluation focused on a forward-looking assessment of the *likelihood* of sustainability. The verification reports will also provide some indication of the sustainability risks.

For each of the three programmes, evidence for assessing *which factors* affect the *likelihood* of long-term sustainability beyond programme end primarily came from:

- midline findings (especially on how design and implementation affected prospects for sustainability);
- verified outcome data and findings of other relevant surveys commissioned during the outcome phase;
- programme documentation, including suppliers' outcome-phase strategies/plans and progress reports;
- findings of country visits (case study countries); and
- interviews with programme managers and verifiers (countries not visited).

For the case study countries, the evaluation team compared the perspectives from different stakeholder interviews, including supplier staff; government partners; lead and country verifiers; and other informants not directly involved in the WASH Results Programme (such as other development agencies active in the sector). This facilitated the ability to triangulate across different perspectives and minimise any biases.

A risk-based framework<sup>25</sup> was used to assess the likelihood of long-term sustainability, by answering the question: *Which presently observable factors imply risks to the sustainability of services, and how large are those risks?*

<sup>25</sup> During 2017, two sustainability workshops were held, both attended by DFID, the suppliers, and verifiers. Amongst other things, these workshops began developing a set of sustainability indicators that could be used for assessing whether an enabling environment for sustainability was in place. The indicator set was not completed prior to the evaluation design, however, and could not therefore be adopted for use in this evaluation. Having said this, risks to sustainability and enabling conditions for sustainability are two sides of the same coin, and the risk factors listed in Table 2 are well aligned with the draft indicators emerging from the workshops.

Table 4 presents the list of prerequisites in an analytical framework which permitted assessment of the extent to which the enabling environment for sustainable service provision has been adequately addressed, as well as what residual risks remain.

The framework also includes enabling environment factors at national level. However, the evaluation also took into consideration the fact that programme design did not include any objectives relating to the policy and institutional framework. Any significant challenges or gaps identified at this level would not, therefore, signify a failure on the part of suppliers. Rather, they would highlight the need for government and development partners to address these issues as a strategic challenge for the sector as a whole.

The table was completed for country case studies for each lead supplier. Both SAWRP (Plan and WaterAid) and SSH4A (SNV) focused on sanitation and hygiene promotion. For SWIFT in Kenya, it was necessary to complete a table for each sub-programme since they differed widely in terms of technical content and context. A summary was then produced of the most important issues and challenges to sustainability emerging from the programme as a whole.

The evaluation team had insufficient information to complete this table for countries not selected for site visits.

**Table 4: A framework for assessing risk to sustainability of WASH services**

Aspect <sup>26</sup>	Areas of investigation	Likelihood			Impact			Risk <sup>27</sup>		
	<i>[Numbers in this row are scores based on professional judgement]</i>	L (1)	M (2)	H (3)	L (1)	M (2)	H (3)	L (1-2)	M (3-4)	High (6-9)
<b>User / community level</b>										
Functional	1. Are the selected technologies and systems fit-for-purpose and fit-for-context?									
	2. Is the construction quality of physical infrastructure adequate?									
Institutional	3. Are the responsibilities of service users and support organisations clearly and appropriately established?									
	4. Are service users organised, trained, and equipped to undertake management tasks for which they are competent and capable?									
	5. Do service users have the means and mechanisms to report faults and request technical assistance?									
Behavioural	6. Has the programme achieved its outcome-level targets? (latrine use; adoption of handwashing with soap; and (where relevant) consumption of safe water)?									
	7. Has there been substantive action during the outcome phase to consolidate latrine use and the adoption of handwashing with soap?									
Financial	8. Did service users make a substantial capital cost contribution? (For household sanitation, this should be the full capital cost, barring cases of exceptional hardship).									
	9. Is there real demand for the services developed, demonstrated through use and payment of operating / repair / replacement costs?									
	10. Will funds collected meet the full lifecycle costs? If not, are arrangements in place for the shortfall to be met by local government or another permanent organisation?									
Environmental	11. Has the long-term adequacy of the quality and quantity of water resources been assessed and, if necessary, addressed? (Including the possible impact of sanitation).									
	12. Have the potential impacts of climate change been assessed and addressed in technology choice and system design?									

<sup>26</sup> Based on the five 'building blocks' of sustainability presented by DFID at recent WASH Results Programme sustainability workshops, plus a sixth: behavioural.

<sup>27</sup> Risk is calculated as the product of the likelihood and consequence scores.

Equity	13. Have the prerequisites for achieving inclusive WASH outcomes been addressed by suppliers?									
<b>Local government level</b>										
Institutional	14. Is external support and guidance (from local government and/or private sector) accessible and responsive to service users' needs?									
	15. In the case of emergencies (e.g. floods), does local government have response arrangements in place to restore services as promptly as possible?									
	16. Do local governments maintain accurate registers of physical assets within their administrative areas, and are asset management plans in place?									
Financial	17. Are goods (e.g. spare parts, sanitary hardware) and support services affordable to service users?									
<b>National level</b>										
Institutional	18. Are sustainability commitments and actions incorporated into sector strategy?									
	19. Is there clarity on the monitoring, management, and financing responsibilities of service users, government (each tier), NGOs, donors, and the private sector?									
	20. Are sufficient funds transferred from national to local government to enable community support and the active monitoring of WASH services?									
	21. Where necessary, are adequate measures in place to develop the capacity of government agencies to play an effective role in service delivery or community support?									
	22. Is a viable sector monitoring system in place or under development?									
	23. Are measures in place to facilitate learning on sustainability, and the application of that learning?									

## **B.4 Endline evaluation activities**

The findings of this report draw on analysis of programme documentation from the earliest days of the design phase to its conclusion: verification data; ongoing analysis of literature on PbR and WASH practice; two phases of interviews with key informants among all stakeholder groups; and, in particular, the findings from purposively selected case studies undertaken for each of the three suppliers. Further information on each activity is included below.

### **Literature review**

The literature review of best practice in the application of PbR which was undertaken in 2015 was repeated in the summer of 2018, to assess if there was any new secondary data to feed into the evaluation. The review not only covered PbR within the WASH sector, but also within the broader context of donor-funded programmes implemented by INGOs.

### **Interviews and roundtable with DFID staff**

Interviews with DFID were held repeatedly over the course of the evaluation, as DFID is a key stakeholder which has a clear overview of the factors affecting the WASH Results Programme design and implementation, as well as the extent to which learning from the WASH Results Programme has been captured and is contributing to better evidence-based WASH policy and programming within DFID. In June 2018 a roundtable discussion was also held with DFID staff to gather feedback on the draft findings and lessons from the evaluation for future PbR programming.

### **Interviews and roundtable with lead suppliers**

The lead suppliers are key stakeholders for the evaluation as they were the main source of evidence about actual supplier behaviour and the factors affecting WASH Results Programme implementation. They were interviewed within the context of planning the case study visits (one-to-one), as well as after the case studies, to fill any gaps identified and to gain information on any countries not visited as part of the case studies.

In June 2018, a roundtable discussion was also held with all key supplier staff to explore emerging findings and lessons learned from programme experience overall across the sub-programmes.

One staff member from one or two consortium members (e.g. Plan and WaterAid) was interviewed in each country so as to obtain multiple perspectives, in addition to the face-to-face interviews carried out with supplier staff in the case study countries during the Q1 2018 country visits.

Remote interviews were also conducted with all country managers (supplier staff). These interviews explored if and why factors affecting implementation identified during the country visits also affected implementation in the remaining countries. We explored barriers and facilitators to help to explain variations in performance between countries.

## **Review of the verification reports**

The six-monthly verification reports produced over the course of the outcome phase, and the associated supplier documentation, were reviewed not only to confirm the quantum of results achieved against targets (DEQ 2.1), but also to assess trends in activity for each sub-programme as well as over the whole 11-country programme.

## **Interviews with lead verifiers**

One-to-one discussions were conducted with lead verifiers to:

- obtain their second-hand perspective on which factors affected the effectiveness, efficiency, impact, and sustainability of the WASH Results Programme, to complement the first-hand perspectives obtained by programme implementers (suppliers) and the programme manager (DFID); and
- draw on the verifiers' first-hand perspective on how the verification system affected effectiveness (HEQ.2.2) and efficiency (HEQ.3.3). These discussions gave us key insights into the learnings from designing and implementing a PbR verification system.

In June 2018, a roundtable discussion was also held with the whole verification team to explore emerging findings and lessons learned from programme experience overall across the sub-programmes.

In addition, the evaluation team also joined the internal learning workshop of the verification team (January 2018) and continued their more informal exchanges of impressions, as necessary.

## **Interviews with Itad's L&D team**

Interviews with Itad L&D team were also carried out to help answer HEQ.5.3, namely understanding what learning has taken place during the programme. The discussion also discussed overlaps and similarities in, as well as differences between, the lessons that came out of our programme evaluation with the lessons the L&D team have drawn from their activities.

## **In-country case studies**

In Q1 2018, visits to case-study countries focused on the outcome-phase implementation activity, focusing in particular on potential impact and the likelihood of

long-term sustainability. The interview guides were tailored to country-specific aspects related to each evaluation theme.

A longer two-week visit took place in Pakistan (SAWRP) and Kenya (SWIFT), and a six-day visit took place in Uganda (SSH4A). In addition, a shorter three-day visit was undertaken in an additional SNV country (Tanzania).

During the country visits, we interviewed supplier staff at country level as they were the main source of evidence about actual supplier behaviour in country. These interviews helped enable the evaluation to more deeply address DEQs relating to:

- whether assumptions in the theory of change (and about PbR generally) held true – for example, whether PbR resolved misaligned incentives and encouraged suppliers to adopt innovative, cost-effective approaches;
- to what extent the short programme timeframe, high targets, and risk of non-payment affected implementation, particularly looking at factors promoting or hindering the sustainable use of sanitation facilities and the long-term adoption of hygienic behaviour;
- whether and how the projects targeted the poorest and most vulnerable or under-served communities / households – and if not, why; and
- how operations and performance were affected by the verification system associated with PbR.

We also interviewed key informants at national level not involved in implementing the WASH Results Programme to obtain an outsider's perspective on the programme's achievements.<sup>28</sup> These interviews allowed us to more deeply assess:

- whether country sub-programmes were aligned with, and supportive of, the national WASH strategy and priorities (DEQ 2.6);
- the extent to which government plays an enabling role in relation to the sustainability of facilities and behaviours (DEQ 2.6); and
- which external, country-specific factors affected the achievement of the programme's outputs and outcomes [DEQ 2.6].

The detailed findings from the case study visit are summarised in Annex D.

## Case study selection

Case study countries were selected purposively and as such are not statistically representative of the programme as a whole. However, for two of the suppliers (SWIFT and SAWRP) there were only two countries from which to choose, whilst SSH4A had eight countries in diverse contexts. 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

<sup>28</sup> During the shorter country visit in Tanzania, we only interviewed supplier staff.



effectiveness of the PbR modality under different circumstances to be explored in detail. Due to budgetary constraints, only three case studies were feasible. The following case study countries were selected and agreed with DFID following sign-off of the Evaluation Design Document in 2015 for the midline, and were visited again at endline:

- **Kenya (SWIFT consortium)** was an interesting case study for understanding how WASH activities and the PbR modality were implemented within a consortium that included very diverse partners and implementation activities.
- **Uganda (SSH4A programme)** highlighted the opposite case, allowing the evaluation team to explore how WASH activities and the PbR modality were implemented by a single INGO applying a standard implementation model across multiple countries. For the endline, the team also conducted an abbreviated three-day visit to Tanzania in order to obtain insight into a second SSH4A country, given that SSH4A was implemented across more countries than the other suppliers.
- **Pakistan (SAWRP consortium)** was implemented by three principal partners. Two of these applied a broadly similar approach to sanitation and hygiene promotion, while the third delivered a dedicated school hygiene promotion intervention with a unique methodology.

## B.5 Analysis of evidence

Most of the evidence that fed into the evaluation of the WASH Results Programme were qualitative data arising from the documentation and stakeholder consultations. A robust and systematic approach to collecting, managing, and making sense of the large volume of qualitative data was therefore critical.

Qualitative analysis consisted of two main layers:

- *content analysis* to draw out findings from individual stakeholder consultations or the document review to help identify common content and subject matter; and
- *thematic analysis*: developing descriptive themes from the primary data and the generation of analytical themes to provide greater context and interpretation of the key findings. This may have been at the level of country (for the case study locations), or higher-level themes (e.g. around how the output/outcome split design affected implementation), and helped ensure a more complete picture of the intervention provided and understanding of the contexts in which it operated.

In parallel to data analysis, the evaluation sought to synthesise the qualitative and quantitative evidence, as well as secondary data, to answer the evaluation questions. This facilitated the evaluation team's efforts to triangulate the evidence and validate the findings from different data collection strategies. The evaluation team could be more confident if similar findings and themes were identified from different methods and data sources. Synthesis also facilitated the ability to generalise the findings from across the different pieces of evidence and to provide insight into the potential replicability of (elements of) this type of programme in different contexts and countries.

One challenge to the synthesis of the outputs from these multiple analytical strands was dealing with any contradiction in the findings produced. To address this challenge, a 'strength of evidence' based approach was used, in which different strands of evidence were weighted based on a quality assessment of their reliability and validity to minimise bias. The weighting varied depending on the type of data collection method, the level of stakeholder engagement, and the evaluation question being addressed (whether questions were answered directly or indirectly through available evidence). For example, a question relating to implementation processes would have been better informed through stakeholder interviews rather than monitoring information, thus more weight would be given to the qualitative evidence.

For example, the data gathered from all the data sources listed in Annex G were analysed within one evidence matrix per supplier, then across all suppliers. The data were analysed to identify patterns in the change across the programme, to identify and understand the causal mechanisms that generated the change, and to draw out recommendations based on the evidence. This matrix summarised the main finding and indicated the strength of the evidence (strong / medium / weak) for that finding. This is an essential step when triangulating across several potentially conflicting data sources.

To limit evaluator bias, the analysis drew on evidence reviews conducted by different members of the evaluation team. Cross-team workshops were held at key stages to assist this process and core members cross-reviewed others' findings. There was also an internal peer review process within the evaluation team and with the advisory group so that members less involved in the design or data collection could take a critical approach to the analytical outputs to help ensure the findings were defensible and could stand up to scrutiny.

## B.6 Evaluation risks and mitigation strategies

The evaluation risks and their level of significance are outlined in the first two columns, with the third column explaining how they were addressed. The last column gives an assessment of the residual risk after mitigation measures were taken. Note that we perceive that none of the risks were high, and only one was viewed as a medium-level risk. The Team Leader of the evaluation and the Evaluation Manager monitored these risks over the course of the implementation period, and took steps to implement the mitigation measures on an ongoing basis, particularly for the highest risks.

**Table 5: Risks and mitigation strategies**

Description of risk	Level of significance	Mitigation measures	Residual risk
<b>Breadth of activities across countries</b> – the WASH Results Programme entails a wide range of implementation approaches. It was a challenge for the evaluation to obtain a	Not being able to evaluate all the implementation approaches of the WASH Results Programme could limit the generalisability of the case study findings.	For all suppliers, the evaluation team increased the effort spent on interviewing country managers in countries that were not visited.  The evaluation team undertook case studies in	<b>Medium</b>

complete picture across all of SSH4A (given that only two countries could be visited through case studies) and SWIFT (given the wide range of partners).		two countries of SSH4A, in contrast to other consortia where only one case study was conducted for each one. Interviews with country managers were partly able to cover other countries.  For SWIFT, efforts were made to clarify findings which held for some partners or groups but not for others.	
<b>We may not be able to interview all relevant stakeholders or programme locations during the country visits and some staff will leave in the last months of programme (loss of institutional memory).</b> This was a particular challenge for both SSH4A case studies: the evaluation team could only spend a brief time in each country, allowing limited opportunity to explore programme-specific issues in depth.	Not being able to include the perspectives of certain in-country stakeholders could limit our assessment of the effectiveness and sustainability of the WASH Results Programme. This could limit the generalisability of the case study findings for all of Uganda or all of Tanzania, given that Mwanza and West Nile faced very different implementation challenges <sup>29</sup> .	All country visits were scheduled for Q1 2018 (the implementation phase ended in March) to reduce the risk of staff having already left their posts.  In hindsight, all supplier staff from Ghana and South Sudan had already left their posts at the time of the interviews. In other countries, staff turnover meant some staff interviewed at midline had left by endline.	<b>Low</b>
<b>Programme implementers and DFID may have contradictory views on key aspects of WASH Results Programme design and implementation.</b>	Obtaining a balanced perspective across stakeholders is important in order to draw rigorous conclusions for the evaluation.	We interviewed a wide variety of stakeholders, including service users and neutral observers not involved in the WASH Results Programme. This helped us obtain a more balanced perspective.  The rigorous approach underlying contribution analysis facilitated our ability to objectively weigh the validity of different perspectives against one another.	<b>Low</b>
<b>Suppliers are not willing / able to provide the financial data necessary to</b>	In their contracts, suppliers agreed to only provide minimal financial reporting. These data do not	As agreed with DFID, more focus was placed on other qualitative aspects of efficiency (one new	<b>Low</b>

<sup>29</sup> It was not possible to visit some programme areas for logistical reasons (i.e. West Nile in Uganda and Mwanza in Tanzania). Mwanza is harder to access, has high malaria rates, weaker governance, higher poverty, and worse roads than Arusha. West Nile is less challenging than Rwenzori, but has a larger refugee population.

carry out a full VFM analysis.	provide sufficient evidence for VFM analysis.	evaluation question was added).	
<b>Limited direct engagement with beneficiaries.</b> The evaluation design meant that feedback from beneficiaries could not be obtained in a representative manner.	This affected the ability of the evaluation to comment on unintended impacts on beneficiary communities.	Where possible, the beneficiary feedback mechanisms established by suppliers were used to collate feedback.	<b>Medium</b>

## B.7 Evaluation limitations and potential for bias

The evaluability assessment undertaken in 2015 addressed how the boundaries of the overall evaluation were set, so as to only include those evaluation aspects which could feasibly be addressed. This assessment was reviewed as relevant for the evaluation questions to be assessed at endline to ensure an appropriate scope. Importantly, comparing the experiences of the WASH Results Programme with a comparable counterfactual, namely a non-PbR WASH programme of similar scale, was outside of the scope of the evaluation.

In addition, there are limitations that relate to the scope of the evaluation.

### Limitations

There were a number of limitations to endline data collection, as summarised in Table 6. Some were supplier- or country-specific, such as the challenges faced in collecting data in Pakistan, while others were relevant for all suppliers, such as the limited insight into financial data and the limited direct engagement with beneficiaries.

One of the biggest limitations of the evaluation related to isolating how the PbR modality affected programme performance from the effect of other programme and non-programme factors. While we designed an approach that aimed 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 was more challenging to compare non-programme views for countries that were not part of the case studies. To mitigate any potential bias, we strove to clearly present who we interviewed and to state any limitations of the findings in all case write-ups.

**Table 6: Limitation and their implications**

Limitation/issue	Implications and mitigating action
<b>Cross-cutting</b>	
<b>Limited direct engagement with beneficiaries –</b> The evaluation design meant that limited third-party data were collected	As part of addressing this limitation the evaluation team explored using beneficiary feedback mechanisms established by suppliers, as feasible, to supplement in-depth interviews conducted with implementation staff: SAWRP: no viable options were identified. SWIFT: no viable options were identified. SSH4A: feedback collected from beneficiaries by SNV as part FGDs held during the sustainability assessments was reviewed by the evaluation.
<b>Limited view of financial data –</b> Due to the commercial and PbR nature of the contracts partners were unwilling to share financial information with the evaluation team. For example, Plan (SAWRP) explicitly requested the following contractual clause, which was subsequently included in the contract: <i>'The supplier will not be required to report to DFID on expenditures on the project. Where DFID or its MV&amp;E contractor wishes to carry out value for money analysis, expenditure data may be requested for this purpose within defined and agreed terms of reference. DFID will not be entitled to carry out an audit of the project based on a schedule of expenditure but may undertake investigations into fraud, bribery and corruption if it so chooses and expect the full co-operation of the supplier.'</i>	This restriction severely hampers the scope for commenting on VFM beyond that represented by the 'prices' paid by DFID as set out in the contracts. The analysis and discussion of efficiency and VFM aspects is based solely on the qualitative reflections of the supplier staff and discussed in relation to contract value as a whole.
<b>Only some countries visited –</b> Due to scope limitations, the review team only visited a subset of target countries. For countries not visited, this part of the evaluation was limited to a review of documents, remote interviews with country programme managers, and interviews with consortium managers and the lead verifier in the UK.	This is a more serious limitation as the views of implementing partner and government staff are not directly captured by the evaluation. This limitation is a feature of the design. In the analysis, careful attention is paid to clarifying where the Bangladesh findings reveal significant similarities or differences between the two country programmes.
<b>County programmes ending prior to evaluation data collection –</b> For SSH4A, activities in South Sudan halted in 2016, and staff in Ghana left their posts in March 2018 (before the evaluation was able to carry out any phone interviews). For SWIFT, the DRC country manager changed posts before the evaluation was scheduled to hold interviews. For SAWRP, some field staff were out of contract by the time the evaluation was scheduled to hold interviews.	For SSH4A, it was not possible to speak to the country managers of Ghana and South Sudan as these were out of contract by the time the endline evaluation activities took place. As such the evaluation is less able to comment on the applicability of the findings in those two countries. Throughout the report this is re-stated where relevant. For SWIFT, the evaluation was able to interview previous managers in DRC even though these were no longer in post. For SAWRP the unavailability of field staff had no implications as, due to security reasons, the evaluation was not able to hold any interviews with field staff regardless.

Specific to a supplier or country	
<p><b>SAWRP: Restricted movement during Pakistan country visit Q1 2018</b> – In November 2017, Plan International was notified by the Government of Pakistan that its permission to operate in the country was not being renewed and that programme operations had to cease within three months. This created considerable uncertainty as to whether the evaluation mission could go ahead. Plan International lodged an appeal and eventually received an indication that they might be allowed to continue, but by the time the mission was confirmed it was too late to obtain the no-objection certificates, which were needed from government so that the evaluation team could visit implementing partner offices and beneficiary communities. No-objection certificates are required for all international staff conducting work in Pakistan if leaving the major cities (Islamabad, Lahore, and Karachi).</p>	<p>The restrictions meant that the evaluation team was unable to leave the major cities, and was unable to visit implementing partner offices or targeted communities.</p> <p>In agreement with Plan International and WaterAid, the mitigating action was for the key informants to travel to meet the evaluation team in the cities. This enabled the team to meet all key implementing partners and a number of government counterparts, though it was generally limited to managerial staff; only a few field staff attended.</p> <p>As most staff were not retained after mid-December 2018, in practice this had a narrower effect than might have been the case if site visits were conducted earlier.</p> <p>The evaluation team had freedom of movement in Pakistan during the midline evaluation.</p>
<p><b>SAWRP: Outcome data availability at the time of the case study</b> – At the time of the mission the final verified results for the outcome phase were not yet available; neither were the results of SAWRP's final SAF surveys.</p>	<p>The implication for data quality is that at the time of interviewing programme staff the final outcomes were not available and therefore could not be considered in interviewing supplier staff. However, as the midline and endline outcome survey results were very similar, with hindsight this is not a significant issue. That is, the final outcome survey results did not raise issues that would invalidate the findings from interviews or necessitate significant further interviews.</p>
<p><b>SWIFT: Staff turnover between phases</b> – The SWIFT Kenya programme experienced a number of staff changes during the outcome phase. In particular, the Programme Coordinator and Monitoring, Evaluation, and Learning Officer both left during the final six months of the programme and the new incumbents were unable to provide an overview of activities and associated issues, challenges and lessons learned over the full course of the outcome phase. There was more continuity, however, within the various sub-programmes.</p>	<p>Findings from midline interviews conducted earlier with programme managers enabled the team to further contextualise the findings from endline interviews. Furthermore, the global manager had previously been involved in the DRC SWIFT programme.</p>
<p><b>SWIFT: Range and variety of projects in Kenya</b> – Given the number and variety of sub-programmes making up SWIFT in Kenya, the evaluation team could only spend a brief amount of time with each one, allowing limited opportunity to explore project-specific issues in depth.</p>	<p>The team were able to engage with staff from all six of the partners working in Kenya at midline and endline. For reporting, the evaluation team has aggregated the results as feasible, and left them disaggregated as necessary. Similarly, care has been taken to clarify findings which hold for some partners or groups but not for others.</p>
<p><b>SSH4A: Breadth of programme across countries</b> – As SSH4A spanned nine countries (with activities in South Sudan discontinued in 2016) it was a particular challenge for the evaluation to obtain a complete picture across all</p>	<p>The evaluation team undertook case studies in two countries of SSH4A. Interviews with country managers provided partial information for the other countries, in key areas, facilitating cross-comparisons.</p>



of SSH4A, given that only two countries could be visited through case studies.	However, the evaluators are less able generalise some findings for all of SSH4A. This is flagged in the report where relevant.
<b>SSH4A: Time constraints in country compared to other country visits</b> – During the two country visits, the evaluation team could only spend a brief amount of time in each country, allowing limited opportunity to explore programme-specific issues in depth. Time constraints meant all the interviews were done over five days in Uganda and in three days in Tanzania. It was not possible to visit some programme areas for logistical reasons (i.e. West Nile in Uganda and Mwanza in Tanzania). In view of the time constraints during the two country visits, it was not possible to meet certain stakeholder groups <sup>30</sup> .	<p>Efforts were made to gather insights from SNV Core Team to understand the activities, successes, challenges, and lessons learned over the full course of the programme.</p> <p>As such, the evaluation is less able generalise some findings for all of Uganda or all of Tanzania, given that Mwanza and West Nile faced very different implementation challenges<sup>31</sup>. This is flagged in the report where relevant.</p>

## Potential bias

There are several sources of potential bias arising from the data collection. In all known cases of possible bias, mitigating actions were taken where possible, as noted below:

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

<sup>30</sup> For example, it was not possible to interview community leaders or Sub-County Health Assistants or Village Health Teams in Uganda or LCBs whose contracts had expired in Uganda and or been discontinued in Tanzania.

<sup>31</sup> For example, Mwanza is harder to access, has high malaria rates, weaker governance, higher poverty, and worse roads than Arusha. West Nile is less challenging than Rwenzori, but has a larger refugee population.



significantly, the verification and the evaluation were independently managed workstreams.

- **All stakeholders consulted, as well as the evaluators themselves, brought their own bias to the evaluation** – Bias in data derived from any human source is inevitable. For example, in this evaluation, every stakeholder's view on the programme's performance was influenced by that stakeholder's own experience and expectations. For this reason, and to support robust analysis, all interviewers were asked to consider the credibility of the interviewee and factors influencing their responses.

To counter any bias within the evaluation team itself (i.e. based on past experiences and expectations), we held regular internal meetings and involved all team members in the development of conclusions and recommendations, which were also independently reviewed by the advisory panel.

- **By its nature, this is a theory-based (generative causal) evaluation, i.e. it is lacking 'counterfactual' evidence.** Considering a counterfactual would have been ideal for the evaluators in answering the evaluation questions. However, it was not possible within the scope of this evaluation or the circumstances of programme to conduct a full counterfactual analysis. Therefore, it is not possible to observe 'what has happened *with* the programme' to compare it to a 'without the intervention' scenario, experimental or modelled.

The evaluation team took several steps to generate a strong body of evidence and reduce bias in the evaluation, and we have consciously taken bias and representation into account in analysing our findings and developing conclusions. For example, we created the evidence framework prior to beginning data collection so as to target particular types and volumes of evidence. We specifically set out to collect the same information from different stakeholder sources to enable triangulation to the extent feasible. Also, we analysed our primary data reported from stakeholders in a critical way so we could weigh up how credible the evidence was. The evidence framework was set up in such a way as to give different 'strengths' to evidence derived from different stakeholders / sources, depending on the evaluation question / hypothesis. Nonetheless, we recognise that for some of the evaluation questions, the evidence gathered may be weaker. Where this is the case we have aimed to be transparent about this in the text and the completed evidence framework in Annex F outlines this further.

## External validity/generalisability

The evaluation aimed to pull together lessons on WASH implementation and PbR design that are relevant to the wide variety of contexts present across the 11 country programmes of the WASH Results Programme. While we designed a rigorous approach for comparing the findings of the case study countries to experiences across the remaining countries, we are aware that not all findings from the case studies are generalisable to the overall programme. When summarising findings from the case studies, we sought to clearly indicate which of these are generalisable, and which are interesting but not widely applicable across the entire programme. Budget constraints limited us from undertaking more country case studies than those that were conducted.

However, we allocated additional resources for interviews with the suppliers in the eight remaining countries at endline to help mitigate this dynamic.

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

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

## B.8 Extent of attribution of results

In the absence of a control group, the evaluation is unable to clearly establish which results can be attributed solely to DFID's funding of the WASH Results Programme and which aspects other funding sources also played a role in. However, there are strong indications that for several country programmes other resources were explicitly leveraged, or accidentally contributed to the achievement of results. This means that the achievements of the WASH Results Programme cannot be solely attributed to DFID funding:

- In those cases<sup>32</sup> where some or all programme activities were implemented directly through government, these government partners contributed substantially in terms of staff time and operational costs. However, these costs were not fully covered by DFID funding<sup>33</sup>.
- In regard to cases where local partners carried out implementation activities, there were several cases<sup>34</sup> where implementing partners reported drawing at times on other funds to support implementation. Though these contributions could not be quantified, these are likely to be smaller than the government contributions mentioned above.

<sup>32</sup> This refers to most SSH4A countries, as well as SAWRP in Bangladesh and SWIFT in DRC, and the work with water utilities in Kenya (Nairobi City Water and Sewerage Company, LOWASCO, and KAWASEPRO).

<sup>33</sup> For example, SNV did not cover government staff salaries, and only covered those facilitation and training costs for government partners which were 'outside of normal operations'.

<sup>34</sup> Several instances were mentioned for SAWRP partners in Pakistan. For SSH4A, two instances of LCBs contributing their own funds are known. This aspect could not be explored in other SSH4A countries. There are no known instances for SWIFT.

- Especially for CLTS activities, a large network of community-level volunteers were involved in promotion activities and in ODF certification. Their costs are not reflected in programme expenditure or the prices per beneficiary figures.

In the majority of programme locations<sup>35</sup>, other government WASH initiatives were being implemented, which created a more favourable enabling environment within which the WASH Results Programme operated. Both national sanitation and hygiene campaigns, as well as local government efforts, will have influenced achievements. This factor particularly played a role in reported results, where results were calculated using population-wide sampling without comparing to a control group<sup>36</sup>. This was the case for all SSH4A countries, and for SWIFT in the ASAL region of Kenya, and for SWIFT in DRC.

## B.9 Dissemination activities

The dissemination activities of e-Pact are distinct for the MV component and the evaluation component, though they are linked at crucial points. The subsections below outline the target audience and objectives of the dissemination activities, summarise what has been done to date, and what activities are planned for the endline so far.

### Target audience of the evaluation

As DFID and the suppliers are the main stakeholders of the evaluation, they are also the main target audience for the evaluation findings. In addition, findings on how to design and implement future PbR programmes in the WASH sector effectively will also be disseminated to WASH practitioners and the wider international community.

### Dissemination aims

The aims of the dissemination activities are as follows:

- To make the evaluation outputs available to DFID, suppliers, WASH practitioners, and the wider international development community.
- To encourage use by DFID and other donors of the evaluation's lessons for applying PbR to WASH programmes.

### What has been done during the midline phase

Several dissemination activities were conducted after the midline:

<sup>35</sup> This aspect was noted for all SSH4A countries, and for DRC, where there is significant donor investment (including non-SWIFT DFID support) in the government VEA programme.

<sup>36</sup> This means that changes over baseline levels would also have been influenced by other WASH initiatives.

- The preliminary evaluation findings relating to programme design were presented and discussed at the 2015 WASH Results Learning Event and shared through the subsequent internal event report.
- In March 2017, the Evaluation Team's Reading List on PbR was published on the MVE Team blog<sup>37</sup>.
- A webinar was used in June 2017 to share the midline evaluation findings and was featured in a Centre for Global Development email newsletter.
- The midline evaluation report and webinar slides are hosted on the OPM website and a blog promoting them was published. These materials were also later uploaded with an accompanying blog post on the MVE team's blog.<sup>38</sup>

## What is planned for endline dissemination

We are in the process of developing an updated communication and dissemination strategy, in consultation with DFID that is expected to be completed after this evaluation is finalised. For example, we are exploring whether additional learning-specific outputs would be helpful to supporting learning by suppliers and DFID. If so, we may combine these notes with tailored workshops for DFID, for suppliers, and for future interested bidders, for example.

So far, we have conducted or are planning the following activities:

- UK-based dissemination workshop (in partnership with the broader e-Pact L&D team) – earmarked for Q1 2019.
- Participation of evaluation team in event at World Water Week with Itad (September 2018).

Further dissemination activities for the endline will likely include:

- making the final reports available on the DFID website;
- participation of evaluation team in evaluation-specific sessions at WEDC 2019 (pending acceptance of abstract);
- a webinar (open to external audiences);
- shorter report summaries and/or policy papers;
- posts on the MVE blog<sup>39</sup>(e.g. reflective posts from the evaluation team, sharing the final report, announcing webinars and other events, etc.)

<sup>37</sup> <https://washresultsmve.wordpress.com/2017/03/16/a-short-reading-list-on-payment-by-results-from-our-evaluation-team/>

<sup>38</sup> <https://washresultsmve.wordpress.com/2017/12/13/wash-results-evaluation-highlights-differences-in-design/>

- posts on consortia partners<sup>40, 41, 42, 43</sup> (e.g. reflective posts and informative posts on WASH interventions); and
- tracking relevant discussion lists and posting information on the evaluation outputs and events, e.g. promoting the webinar on <http://forum.susana.org/>, and sharing results via email lists such as 'sanitation updates'.

## B.10 Governance

The governance structure of the evaluation is based on two processes:

- **The Technical Advisory Committee**, consisting of DFID and the three lead suppliers, will review and sign off on the Evaluation Design Document.
- **The MOU** between DFID, the suppliers, and the MVE provider (e-Pact) complements the formal contractual relationships that DFID holds with the suppliers and MVE. It is set out in the verification volume of the inception report, and makes clear the general principles and processes governing the relationship between DFID, the suppliers, and MVE; the MOU ensures by its inclusive nature that diverse perspectives are taken into account in the evaluation.

The specific endline evaluation scope and activities were overseen by DFID staff with expertise in WASH. The Endline Design Note was reviewed and approved by DFID before data collection began. Suppliers were also provided with an opportunity to review the Endline Design Note in advance of data collection. The Evaluation Manager provided DFID with regular updates on the evaluation activities, including an opportunity to review emerging findings before the team began drafting the final report. DFID and all three suppliers reviewed draft versions of this report.

## B.11 Quality assurance process

We used the following multi-layered quality assurance process to ensure the robustness and integrity of our results:

1. The **evaluation manager** reviewed and edited all outputs, to ensure that they aligned with the TOR, were in accordance with the workplan, and were within budget.
2. The **advisory team**, consisting of WASH and PbR experts, reviewed all key outputs and provided ongoing expert advice on the evaluation design and implementation.
3. The **quality assurance expert** (Alex Hurrell) peer-reviewed the final Synthesis Evaluation Report, in consultation with the subject specialists.

**In addition, specific quality assurance measures were put in place.** A particular challenge for this evaluation was to ensure the quality and consistency of analysis

<sup>40</sup> <https://policy-practice.oxfam.org.uk/our-work/water-sanitation-and-hygiene/swift>

<sup>41</sup> <https://plan-uk.org/about/our-work/healthcare-and-clean-water/clean-water-and-sanitation/sawrp-our-partnership>

<sup>42</sup> <http://www.snv.org/project/sustainable-sanitation-hygiene-all-results-programme>

<sup>43</sup> <http://www.itad.com/introducing-dfid-wash-results/>

across three very different supplier contexts in almost a dozen countries and a large number of stakeholders at programme level and in country. Quality assurance measures included the following:

- Use of a strength of evidence protocol, to allow us to establish an overall confidence level for each finding/conclusion in this report.
- In this report we clearly distinguish between descriptions (e.g. describing the programme implementation) and evaluation findings. The report was structured along the DAC criteria, where possible cross-referencing previous sections rather than repeating evidence.
- We conducted internal review of this report against the EQuALS criteria before submitting it to DFID.
- To the extent feasible given varying contexts, we used standard semi-structured questionnaires and/or recording formats for interviews, meetings, and FGDs/roundtable discussions. These were pre-tested to the extent feasible, first internally and then externally. For example, updates to the initial interview guides were made after the first country visit to facilitate standardisation, as relevant.
- We had regular interaction (informal or structured, remote or in-person) between team members to share and review data and emerging findings.
- All stakeholders were given the opportunity to comment on the draft findings.

## B.12 Ethical conduct

OPM follows strict ethics guidelines in KII, particularly around consent. This included taking verbal consent and ensuring that the participant is clear on how the data would be used (e.g. purpose of the study, procedures for the study, confidentiality and anonymity) and that their participation is voluntary (i.e. that they have the right to refuse).

OPM adhered to DFID's Standards for conduct of ethical research<sup>44</sup> in the following manner:

- 1. Researchers and evaluators are responsible for identifying the need for and securing any necessary ethics approval for the study they are undertaking.** *The independent OPM Ethical Review board granted ethical approval for the research in 2014.*
- 2. Research and evaluation must be relevant and high quality with clear developmental and practical value.** *The evaluation defined its objectives which include generating learning for practical use.*
- 3. Researchers and evaluators should avoid harm to participants in studies.** *Guidelines were followed to ensure no harm was done, including obtaining verbal consent.*
- 4. Participation in research and evaluation should be voluntary and free from external pressure.** *All participants were explained how the data would be used and that their participation was voluntary.*

<sup>44</sup> July 2011, DFID Ethics Principles for Research And Evaluation



5. **Researchers and evaluators should ensure confidentiality of information, privacy and anonymity of study participants.** *It was communicated to all respondents that their responses would be anonymous. Also all sources of evidence were anonymised.*
6. **Researchers and evaluators should operate in accordance with international human rights conventions and covenants to which the United Kingdom is a signatory, regardless of local country standards. They should also take account of local and national laws.** *International standards concerning human rights were adhered to in the conduct of the research and local and national laws complied with.*
7. **DFID funded research and evaluation should respect cultural sensitivities.** *The research team took into account differences in culture, local behaviour and norms, religious beliefs and practices, sexual orientation, gender roles, disability, age and ethnicity and other social differences and were mindful of research burden on different groups.*
8. **DFID is committed to publication and communication of all evaluations and research studies.** *The dissemination plans of the evaluation are set out in Annex B.8.*
9. **Research and evaluation should usually be independent of those implementing an intervention or programme under study.** *Box 6 below provides an assessment of the extent to which the evaluation team was able to operate independently and free from interference.*
10. **All DFID funded research/ evaluation should have particular emphasis on ensuring participation from women and socially excluded.** *Gender and social inclusion aspects were one of the aspects explored in the evaluation.*

#### Box 2: The extent to which the evaluation team operated independently

All stakeholder groups were cooperative and agreed to be interviewed. All three supplier teams, as well as the MV team, were given the opportunity to provide comments on, but not directly edit, a draft version of this report. The team experienced no undue interference from direct stakeholders.

However, as is noted in the limitations section, the team was not able to operate completely independently. Due to circumstances within Pakistan, the evaluation team was unable to travel to meet all stakeholders originally targeted during the site visit. To the extent feasible, the evaluation team followed up via telephone and/or email with the relevant stakeholders. The issue became known just prior to the site visit commencing and modifications to the approach were discussed with DFID in advance.

## B.13 Full evaluation matrix

This annex contains the full evaluation matrix, including our evaluation questions, data sources, and methodological approach and/or basis for evaluation judgement. The last two columns refer to the relevant sections of this report and the confidence level of the findings.

In the few cases where the wording of the evaluation question was updated for the endline, the new wording is provided in this table.

DEQs	Data sources	Methodology / basis for evaluation judgement	Endline section
<b>HEQ1 – relevance:</b> 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 <b>programme objectives</b> clearly articulated? <i>Not revisited at endline</i>	<b>Primary</b> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> <li>Unsuccessful bidder interviews (midline only)</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>DFID documentation</li> <li>WASH best practice review</li> </ul>	<b>Document review:</b> Evidence of clear articulation of all of DFID's underlying programme objectives in the business case, the invitation to tender, and other DFID documentation. This was triangulated with the views of DFID, suppliers, and unsuccessful bidders.  Consistency of stated objectives between documents.	Section 3.1
DEQ 1.2: To what extent does the <b>programme's design</b> (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? <i>Not revisited at endline</i>	<b>Primary</b> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>DFID documentation</li> </ul>	<b>Document review:</b> 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.	Annex C
DEQ 1.3: To what extent were the <b>scale and pace of the programme</b> (including the December 2015 deadline) realistic for achieving the intended outputs and outcomes, given the capacity of suppliers and their local partners? <i>Not revisited at endline</i>	<b>Primary</b> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> <li>Unsuccessful bidder interviews (midline only)</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>DFID documentation</li> </ul>	<b>Document review:</b> Evidence in the invitation to tender of tailoring of WASH Results Programme design to capacity and context, triangulated with the views of DFID, suppliers, and unsuccessful bidders.	Section 4.1.1
DEQ 1.4: To what extent was the <b>PbR modality</b> designed appropriately for achieving the planned results, given the	<b>Primary</b> <ul style="list-style-type: none"> <li>DFID interviews</li> </ul>	<b>Critique of theory of change:</b> Evidence of the expected causal link between PbR incentives and expected theory of change mechanisms, review of incentives acting through the supplier-	Section 3.1; Also refer to Sections 4.4.2 and 4.5.2

capacity of suppliers and the timeline of the programme?	<ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Unsuccessful bidder interviews</li> <li>Lead verifier interviews</li> </ul> <p><b>Secondary</b></p> <ul style="list-style-type: none"> <li>DFID documentation</li> <li>Verification Design Document</li> </ul>	<p>specific verification methodology, triangulated with the views of DFID, suppliers, verifiers, and unsuccessful bidders.</p> <p><b>Process evaluation:</b> Assessment of the ways the varying capacities of the suppliers interacted with the PbR incentives, and if this facilitated or hindered results achievement, through interviews with lead verifiers and suppliers.</p>	
<p>DEQ 1.5: How likely was it that the programme design would encourage 'innovative' private sector partnerships?</p> <p><i>Not revisited at endline</i></p>	<p><b>Primary</b></p> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> <li>Unsuccessful bidder interviews (midline only)</li> </ul> <p><b>Secondary</b></p> <ul style="list-style-type: none"> <li>DFID documentation</li> <li>WASH best practice review</li> </ul>	<p><b>Critique of theory of change:</b> Evidence of clear intent of the programme to encourage 'innovative' private sector partnerships, based on DFID documentation.</p> <p>Assessment of likelihood of private sector partnerships being established, based on triangulation of perceptions of DFID, suppliers, and unsuccessful bidders.</p>	Section 3.1; Also refer to <i>Midline Evaluation Report</i> .
<p>DEQ 1.6: How likely was it that the programme design would encourage suppliers to propose 'innovative WASH interventions'?</p> <p><i>Not revisited at endline</i></p>	<p><b>Primary</b></p> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> <li>Unsuccessful bidder interviews (midline only)</li> </ul> <p><b>Secondary</b></p> <ul style="list-style-type: none"> <li>DFID documentation</li> </ul>	<p><b>Critique of theory of change:</b> Evidence of clear intent of the programme to encourage 'innovative' WASH interventions, inclusive results, and learning, based on DFID documentation.</p> <p>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.</p>	Section 3.1; also refer to Section 4.2.5 or the Midline Evaluation Report.
<p>DEQ 1.7: How likely was it that the programme design would encourage <b>inclusive</b> outputs and outcomes?</p> <p><i>Not revisited at endline</i></p>			Section 3.1; also refer to Section 4.4.2
<p>DEQ 1.8: How appropriate was the WASH Results Programme's design for achieving the programme 'learning objectives'?</p> <p><i>Not revisited at endline</i></p>			Sections 4.2.8 and 4.5.6
<p>DEQ 1.9. To what extent was the <b>design of each consortium sub-programme</b> appropriate for achieving DFID's key objectives?</p> <p><i>Not revisited at endline</i></p>	<p><b>Primary</b></p> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Supplier country case studies</li> <li>Country manager interviews</li> </ul>	<p><b>Critique of theory of change:</b> 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.</p>	Section 4.1

	<b>Secondary</b> <ul style="list-style-type: none"> <li>Quarterly verification reports</li> </ul>		
<b>HEQ2 – effectiveness:</b> 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 <b>intended outputs</b> at scale? <i>Not revisited at endline</i>	<b>Primary</b> <ul style="list-style-type: none"> <li>Supplier country case studies</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>Quarterly verification reports</li> <li>Supplier monitoring data</li> </ul>	<b>Review of verification data:</b> Quantitative assessment of % of output targets achieved in each country by 2016 (disaggregated by implementing partner if possible).  <b>Process evaluation:</b> 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.	Section 4.2.1
DEQ 2.2: To what extent have the prerequisites for achieving <b>inclusive WASH outcome</b> been addressed by suppliers?	<b>Primary</b> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Supplier country case studies</li> <li>Country manager interviews</li> <li>Local partner interviews</li> <li>Community-level FGDs and key informant interviews (KIIs)</li> <li>Pakistan quant/qual research study</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>Quarterly verification reports</li> <li>Supplier monitoring data</li> <li>Supplier baseline/endline data</li> </ul>	<b>Review of verification data and supplier outcome surveys:</b> Quantitative assessment of % of outputs sustained until endline (disaggregated by implementing partner if possible).  <b>Process evaluation:</b> Comparison of the prerequisites for sustainable and equitable WASH outcomes (as set out in Design Document) with what was implemented.	Section 4.4.2
DEQ 2.3: To what extent have services continued to function and have behaviours continued to be used since their initial implementation ( <b>sustainable outcomes</b> )?			Section 4.2.1
DEQ 2.4: How did programme design and external factors affect the achievement of output and outcome objectives within consortia sub-programmes?	<b>Primary</b> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Lead verifier interviews</li> </ul>	<b>Review of verification report:</b> Review of the risk factors cited in the quarterly Verification Reports as affecting implementation in under- and over-performing countries.	Section 4.2 (all)

	<ul style="list-style-type: none"><li>• Supplier country case studies</li><li>• Country manager interviews</li><li>• National KII</li><li>• Local partner interviews</li><li>• Community-level FGDs and KIIs</li></ul>	<b>Contribution analysis:</b> Assessment of the main factors affecting implementation across under- and over-performing countries, based on triangulation of perceptions of DFID, suppliers, and verifiers.	
DEQ 2.5: Under which circumstances did the <b>PbR framework</b> help/hinder the achievement of the intended outputs and outcomes?	<b>Secondary</b> <ul style="list-style-type: none"><li>• Quarterly verification reports</li></ul>	<b>Review of verification report:</b> Review of the PbR effects cited in the quarterly Verification Reports.	Section 4.2 (all)
DEQ 2.6: Under which circumstances did the <b>PbR framework</b> affect the quality of programme implementation (positive or negative)?		<b>Contribution analysis:</b> 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.	Section 4.2.2
DEQ 2.7: Under which circumstances did suppliers implement <b>innovative</b> approaches and focus on learning?		<b>Process evaluation:</b> 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.	Sections 4.2.5, and 4.2.8
<b>HEQ3 – efficiency:</b> Has the programme been designed and implemented in a cost-efficient manner?			
DEQ 3.1: How efficient was the <b>tendering and procurement process</b> and what effect did this have on programme delivery? <i>Not revisited at endline</i>	<b>Primary</b> <ul style="list-style-type: none"><li>• DFID interviews</li><li>• Lead supplier interviews</li><li>• Unsuccessful bidder interviews (midline only)</li></ul>	<b>VFM analysis:</b> Evidence of transaction costs, based on triangulation of perceptions of DFID, suppliers, and unsuccessful bidders.  <b>VFM analysis:</b> Evidence of consequences resulting from transaction costs, based on triangulation of views across supplier staff in country and at HQ.	Section 3, and midline evaluation report
DEQ 3.2: Under which circumstances did the <b>PbR modality</b> affect the cost-effectiveness of individual sub-programmes?	<b>Primary</b> <ul style="list-style-type: none"><li>• Lead supplier interviews</li><li>• Supplier country case studies</li><li>• Country manager interviews</li></ul>	<b>Process evaluation:</b> 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.  <b>VFM analysis:</b> Quantitative assessment of the staff time associated with verification burden in country case studies (not feasible in 2016).	<i>[DEQ removed in consultation with DFID]</i>
DEQ 3.3: Under which circumstances did the <b>PbR modality</b> strengthen the programme monitoring of individual sub-programmes?		<b>Process evaluation:</b> Evidence of examples of how strengthened M&E arrangements led to improved implementation.	4.2.3 and 4.3.6

DEQ 3.4: Under which circumstances did <b>key programme features</b> affect cost-effectiveness?		<b>VFM analysis:</b> VFM assessment comparing the cost and price per beneficiaries and isolating key drivers of cost-effectiveness.	<i>[DEQ removed in consultation with DFID]</i>
DEQ 3.5: How did the efficiency of the <b>programme management arrangements</b> of individual sub-programmes affect programme delivery?	<b>Primary</b> <ul style="list-style-type: none"> <li>DFID interviews</li> <li>Lead supplier interviews</li> <li>Supplier country case studies</li> </ul>	<b>Process evaluation:</b> Evidence of how consortium management arrangements affected implementation, based on triangulation of views across supplier staff in country and at HQ, and DFID.	Section 4.1.1 and 4.3.2
DEQ 3.6: To the extent <b>PbR risk-sharing</b> arrangements were applied within consortia, how did this affect programme delivery?	<b>Primary</b> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Supplier country case studies</li> <li>Country manager interviews</li> </ul>	<b>Process evaluation:</b> Evidence of how the extent of PbR risk-sharing affected implementation, based on triangulation of views across supplier staff in country and at HQ.	Section 4.3.1 and 4.3.3
<b>HEQ4 – impact:</b> 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 <b>impacts</b> ?	<b>Primary</b> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Lead verifier interviews</li> <li>Supplier country case studies</li> <li>Country manager interviews</li> </ul>	<b>Process evaluation:</b> Health assessment against the prerequisites for WASH health impacts (as set out in Design Document).  <b>Impact assessment:</b> Indications of non-health impacts (explanation of causal links / external drivers) uncovered during case studies and through interviews with country managers.	Section 4.4.1
DEQ 4.2: Under which circumstances did the WASH Results Programme activities have any <b>unintended/unplanned positive or negative impacts</b> ?	<ul style="list-style-type: none"> <li>National KII</li> <li>Local partner interviews</li> <li>Community-level FGDs and KIIs</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>Quarterly verification reports</li> <li>WASH sector review for case studies</li> </ul>	<b>Review of Verification Reports:</b> Review of any unintended effects cited in the quarterly Verification Reports.  <b>Process evaluation:</b> Evidence of how implementation factors and external drivers caused unintended impacts.	Section 4.4.3
<b>HEQ5 – sustainability:</b> 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 <b>individual sub-programmes</b> designed and implemented to maximise the likelihood of achieving long-term sustainable WASH outcomes and impacts?	<b>Primary</b> <ul style="list-style-type: none"> <li>Lead supplier interviews</li> <li>Lead verifier interviews</li> <li>Supplier country case studies</li> </ul>	<b>Review of Verification Reports:</b> Review of the risks cited in the sustainability RAG ratings in each Verification Report.  <b>Process evaluation:</b> Assessment against the prerequisites for sustainable WASH outcomes (as set out in Design Document),	Sections 4.5.1 and 4.5.2



	<ul style="list-style-type: none"> <li>Country manager interviews</li> <li>Local partner interviews</li> <li>Community-level FGDs and KIIs</li> </ul> <b>Secondary</b> <ul style="list-style-type: none"> <li>Quarterly verification reports</li> </ul>	based on triangulation of examples given by DFID, suppliers, national sector stakeholders, and beneficiaries.	
DEQ 5.2: Under which circumstances has the <b>PbR modality</b> affected the likelihood of the long-term sustainability of the outcomes and impacts?	<i>As above +</i>	<b>Process evaluation:</b> 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.	Section 4.5.5
DEQ 5.3: Under which circumstances have <b>other programme features</b> affected the likelihood of the long-term sustainability of the outcomes and impacts?	<ul style="list-style-type: none"> <li>National KII</li> </ul>		Section 4.5.1– 4.5.3
DEQ 5.4: Under which circumstances did the WASH Results Programme contribute to enhanced sector learning to inform <b>better evidence-based WASH policy and programming</b> ?	<i>As above +</i> <ul style="list-style-type: none"> <li>Interview with Itad L&amp;D team</li> </ul>	<b>Process evaluation:</b> 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.	Section 4.5.6

## Annex C Critique of the theories of change and associated assumptions

*This annex has been largely reproduced from the midline evaluation report, with only minor updates to the supplier-level impact pathways.*

This annex presents an explicit articulation of the WASH Results Programme's theory of change and its underlying assumptions. These form the basis for both the midline and endline evaluation activities, with the evaluation questions set out in Annex B.2 testing key assumptions underlying the theory of change.

For evaluation purposes, we distinguish between theories of change at three levels:

- Firstly, a 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 C.1 below. This evaluation therefore relies on a non-statistical, theory-based evaluation, and elements of both **realist evaluation** and **contribution analysis** were used, tailored to the specificities of the WASH Results Programme, as indicated below.
- 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 C.2 below. 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.
- Thirdly, the programme design created elements that could be interpreted differently at the supplier sub-programme level, potentially resulting in different impact pathways (process – output – outcome). This is also nested within the programme-wide theory of change and is presented in Section C.1 below.

The distinction made between theories of change at three 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, as well as a supplier-level, perspective.

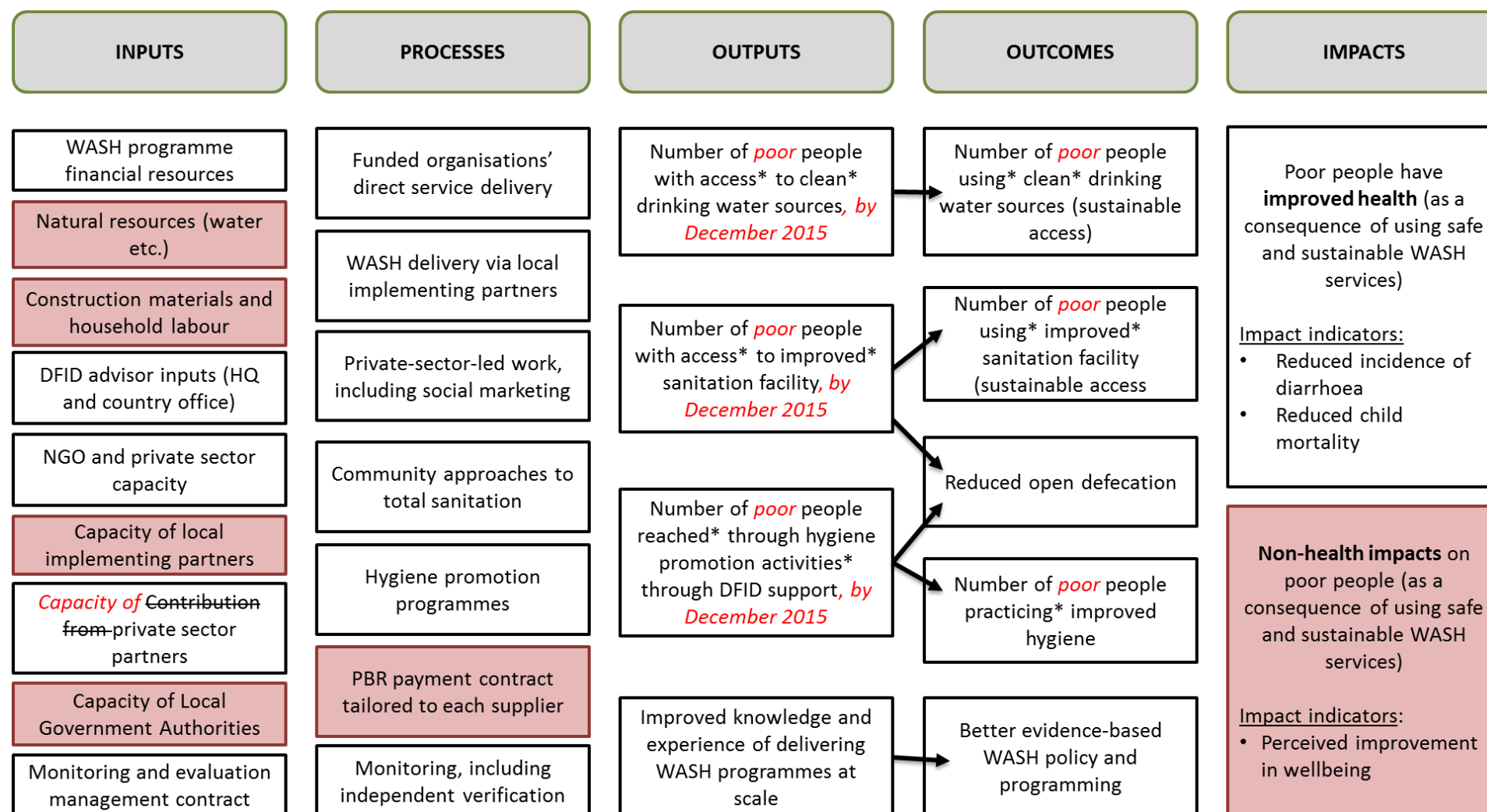
### C.1 The WASH Results Programme's theory of change

Figure 1 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. This 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 C.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 allowed the service provider flexibility in regard to how best to achieve outputs and outcomes given the implementation context.

**Figure 1: 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.

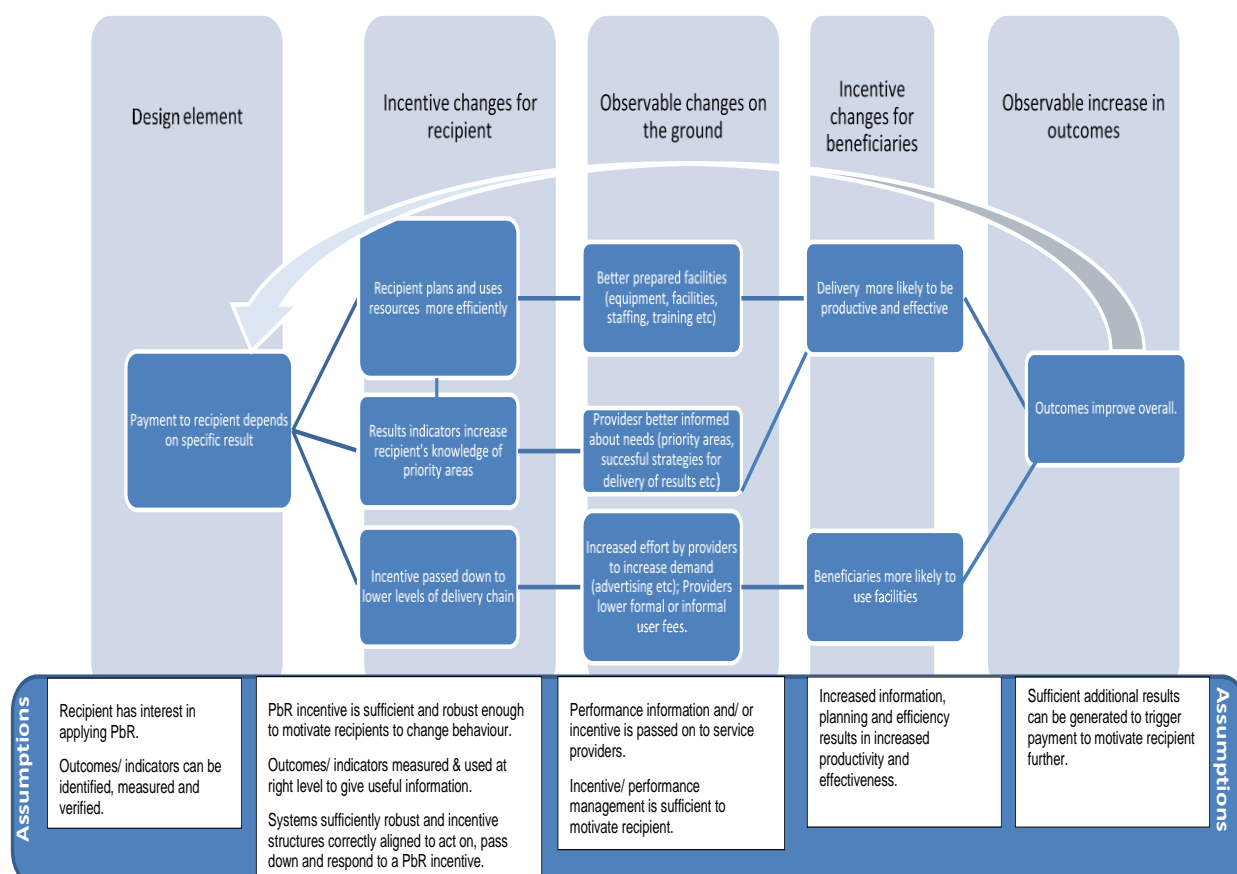
**Table 7: Assumptions stated to be underlying DFID's theory of change**

Inputs to processes	Processes to outputs	Outputs to outcomes	Outcomes to impact
<ul style="list-style-type: none"> <li>Sufficient engagement, level of expertise, continuity, and speed of response by DFID adviser (#53)</li> <li>MVE contract enables 'prompt corrective action' (#124)</li> <li>Sufficient user demand for services (and capacity of the community to manage improved services) (evidence papers)</li> </ul>	<ul style="list-style-type: none"> <li>Any construction and capacity development activities are appropriate and of high quality</li> <li>Vulnerable groups are included in programme (#23)</li> <li>Households make labour and capital available (#37)</li> <li>Private sector has capacity to construct and rehabilitate water-points (#37)</li> <li>Communities are able to operate and maintain the improved water supply (#37)</li> <li>Local government authorities have the capacity to manage inputs and deliver outputs (#37)</li> <li>Hydrogeological, hydrological, and water resources management assessments have been undertaken (#68)</li> <li>Climate risk assessments undertaken (#78, 79, 80)</li> <li>Sustainability assessments undertaken (#82)</li> <li>Effective targeting of poor/vulnerable groups (#117–121)</li> </ul>	<ul style="list-style-type: none"> <li>Households change their behaviours as a result of sanitation and hygiene campaigns (#37)</li> <li>Lifecycle costs and responsibilities for their payment have been included (#15)</li> <li>Community participation and capacity (#15, #16)</li> <li>No impact of time-bound PbR targets on sustainability (#117–121)</li> </ul>	<ul style="list-style-type: none"> <li>A critical mass of households change their behaviours, resulting in health and non-health benefits at community level (evidence papers)</li> <li>No unsustainable abstraction of water or pollution from sanitation (#68)</li> <li>No use of contaminated water sources (#68)</li> <li>No contamination of water after collection from a 'safe' source (evidence papers)</li> </ul>

## C.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 Section C.1, above), complemented by an explicit nested theory of change for the PbR modality (see Figure 2). 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 2: 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:

## Design assumptions

### With regard to inputs and processes

- Service providers' incentives are perceived to be the main constraint to achieving the scale of MDG-related WASH outcomes that the WASH Results Programme aims to achieve.
  - PbR will be an effective tool for shifting the incentive frameworks within service providers (predominantly INGOs).
  - Inefficiencies in INGO practices are perceived to be a key constraint to reducing the unit cost of the MDG-related WASH outcomes that the WASH Results Programme aims to achieve.
  - 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 subcontracted 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, the 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 overachievement, and/or penalisation for underachievement), 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:
  - the verification cycle timeframe is consistent with contractual obligation and suppliers' cash flow constraints; and
  - 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 underachievement 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 suboptimal programme interventions (e.g. less risky).

- The PbR modality does not encourage gaming (e.g. deliberately improving performance measures).

### C.3 Supplier-level impact pathways

Of all the programme features highlighted in Section 3.1, the midline evaluation report listed three intertwined factors as arguably the strongest determinants of how the programme has played out. Their individual effects are tightly intertwined. These are:

- (a) the extremely tight and strict deadline for delivery of beneficiary number results by December 2015;
- (b) the programme 'shape' or phasing, involving output and outcome periods; and
- (c) the PbR modality of payment (including the requirement for rigorous independent verification).

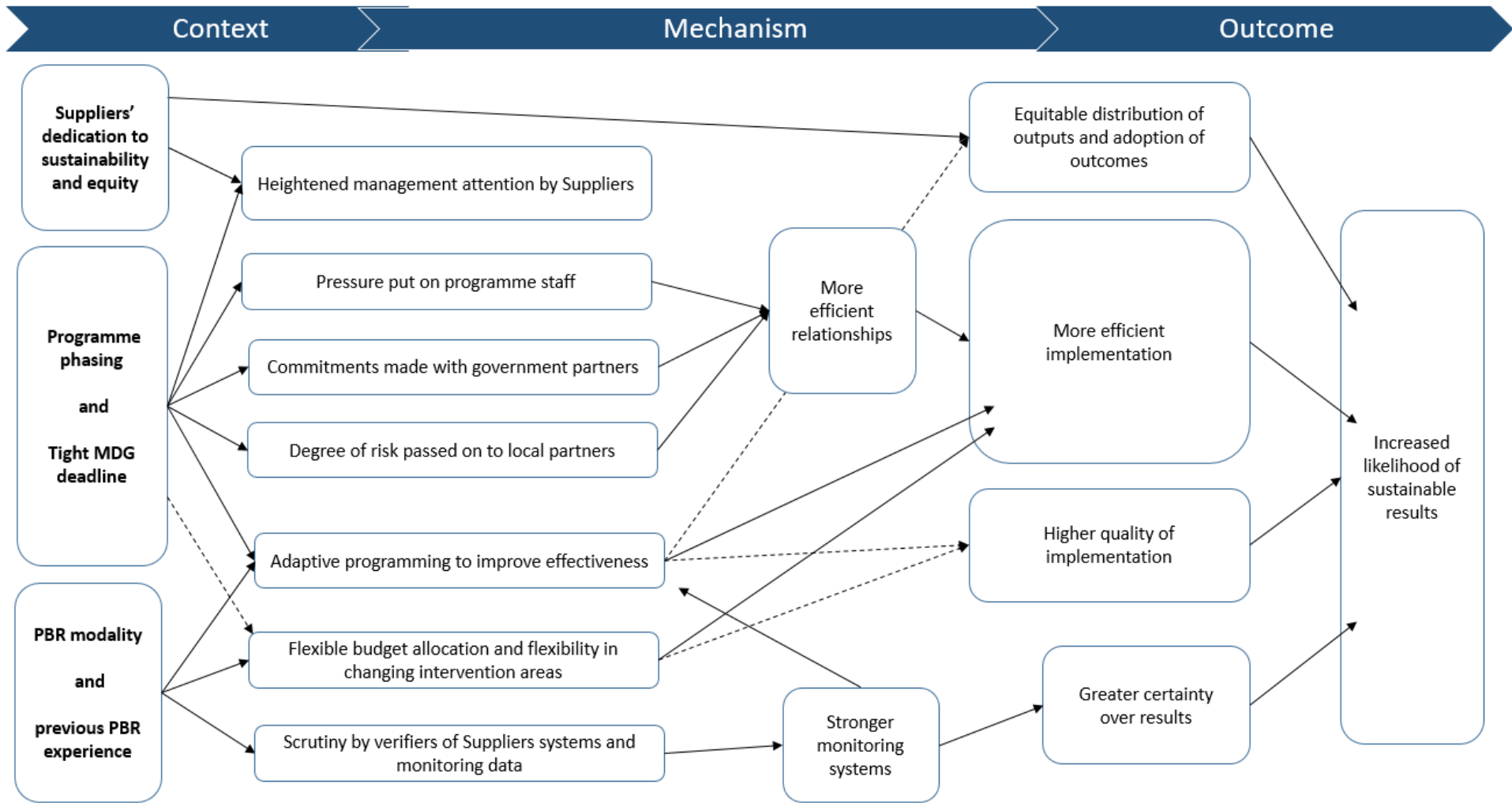
To these factors, two were added during the endline, namely:

- (d) the suppliers' dedication to sustainability and equity; and
- (e) the varying PbR experience of suppliers.

At the supplier sub-programme level, these factors resulted in a number of mechanisms, which influenced efficiency, effectiveness, inclusiveness, and quality in a variety of ways.

The figure below (overleaf) provides a diagrammatic representation of the resulting impact pathways.

Figure 3: Examples of impact pathways



Note that different changes in behaviour could have positive or negative effects on outcomes

## Annex D Summary of contribution analysis

### D.1 Overview of contribution analysis

The most rigorous forms of impact evaluation (such as experimental approaches) address the issue of attribution by identifying a counterfactual. In the case of this evaluation of the WASH Results Programme, it was not possible to identify a counterfactual (i.e. to test what would have happened in the absence of the programme).<sup>45</sup>

This evaluation therefore relies on a non-statistical, theory-based evaluation, and elements of both **realist evaluation** and **contribution analysis** were used, tailored to the specificities of the WASH Results Programme, as indicated below.

#### Box 3: Rationale for contribution analysis in this evaluation

The use of contribution analysis benefits from having the comparison (counterfactual) analysis 'built in' to some degree, because alternative explanations are explicitly explored as part of the analysis. Contribution analysis focuses on understanding the relative impact of a programme. It seeks to differentiate between the impact of a particular intervention (in this case the WASH Results Programme) and other explanations of the outcomes, exploring the 'contribution' of the programme to the observed results.

The approach rests on the creation of a 'contribution story' or 'performance story', providing an initial narrative of what it is reasonable to expect the WASH Results Programme, especially its PbR modality, to have contributed relative to other inputs, as well as 'competing performance stories' or alternative theories and other factors that may be influencing project performance.

Through an iterative process of collecting data that prove or disprove the hypotheses of the contribution story, alternative theories or extraneous factors can be discounted as explanations for the outcomes, and more confidence can be given to our assessment of the WASH Results Programme's contribution to the outcomes.

Contribution analysis was used for DEQs 2.2, 2.4, 2.5, 2.6, 3.4, 5.1 and 5.2 as it helped tease out the factors that contributed to the quality of implementation and achievement of intended outputs and outcomes. The contribution analysis approach considered the wider context, including identifying external factors and mechanisms that may also have been working (or hindering progress) towards the objectives, and sought to pinpoint the WASH Result Programme's own contribution.

The contribution analysis approach was tailored to the needs of this evaluation in several ways:

- First, we not only assessed *if* the WASH Results Programme contributed to the outcome or impact of interest, but also *which elements* of the WASH Results Programme played a significant role in driving this contribution.
- Secondly, we triangulated contribution claims based on information from suppliers and country managers with country case studies and validated them with country-level stakeholders and programme-level stakeholders, adding to the validity of the findings.
- Third, we blended the contribution analysis approach with principles of the realist evaluation approach, particularly at the level of how the WASH Results Programme was having an effect on ultimate beneficiaries. Realist evaluation emphasises under what circumstances and how the programme achieved desired outcomes, embracing the variation in context as well as in

<sup>45</sup> An RCT was conducted under a separate workstream for the WASH Results Programme in Pakistan.

programme implementation to provide an in-depth description of how and why change happened.

- As implementation approaches varied by supplier, the evaluation team probed how variations in the sub-programmes may have altered or influenced the overall contribution story developed. WASH-specific causal mechanisms (i.e. how and why WASH outcomes were achieved given certain conditions) will still be relevant regardless of the programme activities; and, the evaluation could examine whether such mechanisms were ‘triggered’ to produce outcomes through the interaction between programme activities and context.

The following data sources were used in each subsequent phase of the contribution analysis:

	Phase I (midline and documentation)	Phase II (country case study)	Phase III (other country interviews)	Phase IV (roundtables and report comments)
<b>Data sources</b>	<ul style="list-style-type: none"> <li>• Midline evaluation report</li> <li>• Verification reports</li> <li>• Payment decision meeting notes</li> <li>• Other supplier documentation if available (learning reports)</li> </ul>	<ul style="list-style-type: none"> <li>• Information gathered from interviews with various stakeholders while in country – including supplier teams, local partners, and country verifiers</li> </ul>	<ul style="list-style-type: none"> <li>• Remote interviews held with country managers</li> <li>• Interviews held with supplier leads</li> </ul>	<ul style="list-style-type: none"> <li>• Insights from roundtables with DFID, verifiers, and suppliers</li> <li>• Insights from comments made on draft reports</li> </ul>

## D.2 Contribution analysis framework

The contribution analysis framework helped to guide the evaluation’s exploration of outcomes at the supplier level, and helped to guide its assessment of the extent to which the WASH Results Programme specifically can be linked to progress against these outcomes, as opposed to a range of external factors and influences. The contribution analysis framework is presented in Table 8 below.

**Table 8: Contribution analysis framework**

Outcome hypothesis	DEQ	Assumption underpinning TOC	Other contributing factors (external to this programme)	Alternative routes to impact (e.g. other contributing initiative)
H1a. The introduction of a PbR modality helped achieve intended outputs and outcomes	2.4, 2.5, 3.5, 3.6	[See assumptions in nested hypotheses below].	[See other contributing factors in nested hypotheses below]	[See alternative routes in nested hypotheses below]
H1.1a. The programme and its PbR modality allowed flexibility of implementation approach within the sub-programmes, which helped achieve output and outcome objectives	2.4, 2.5, 2.7	<ul style="list-style-type: none"> <li>Assumes that the overall requirements from DFID ultimately allowed flexibility</li> <li>Assumes that suppliers (based on their experience) were able to design suitable WASH activities which deliver agreed outputs and outcomes (within the context of external factors)</li> <li>Assumes that the flexible design gave suppliers increased flexibility, e.g. to adjust their choice of locations, choice of activities, and level of effort (based on their experience and the country context) to ensure that outputs and outcomes were achieved effectively</li> </ul>	<ul style="list-style-type: none"> <li>The national contexts can help or hinder the ability to be flexible or options available, e.g.: <ul style="list-style-type: none"> <li>political momentum behind achieving WASH agenda helps achieve WASH outcomes (e.g. high in Zambia/Ethiopia – low in Mozambique/Ghana)</li> <li>assumes that political/environmental shocks do not hinder operations in country (as they did with Ebola in Liberia, earthquake in Nepal, and insecurity in South Sudan)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Other national WASH programmes can boost WASH services (e.g. national handwashing day in Pakistan)</li> </ul>
H1.2a. Stronger monitoring systems as a result of the PbR modality increased the likelihood of achieving intended outputs and outcomes	2.4, 2.5, 3.3	<ul style="list-style-type: none"> <li>The capacity of the suppliers affected their ability to monitor accurately</li> <li>Assumes that NGOs had poor/inaccurate results data, and that improvement above this baseline was possible and desirable</li> <li>Assumes that the verification process strengthened the monitoring capacity and monitoring systems of suppliers – thus leading to better data</li> <li>The scrutiny of the monitoring data which took place as part of the verification process improved the accuracy/certainty of result achieved</li> <li>Assumes the increased accuracy/certainty led the suppliers to make changes as necessary to increase likelihood of achieving outputs and outcomes</li> </ul>	<ul style="list-style-type: none"> <li>The national contexts can help or hinder the accurate monitoring / surveying of results, e.g.: <ul style="list-style-type: none"> <li>assumes that political/environmental shocks do not hinder operations in country (as they did with Ebola in Liberia, earthquake in Nepal, and insecurity in South Sudan)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>The push for better monitoring data under the SDG agenda can improve certainty over WASH results, separate from a PbR modality</li> <li>Continued learning and best practice in the WASH sector can monitoring</li> </ul>



		<ul style="list-style-type: none"> <li>Assumes that suppliers passed risk down to implementing partners as necessary, to ensure results were achieved</li> <li>Assumes the achievement of intended outputs and outcomes increased the likelihood of sustainable results</li> </ul>		systems over time
H1.3a. The results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes	2.4, 2.5, 2.7	<ul style="list-style-type: none"> <li>Assumes that results-oriented problem-solving is favoured by the PbR modality (compared to other funding modalities)</li> <li>Assumes that results-oriented problem-solving led suppliers to adjust their choice of locations, choice of activities, and level of effort to ensure targets were reached</li> <li>Assumes that the pressure to achieve results left enough time to reflect on progress, and to learn based on which implementation approaches worked well and which worked less well</li> <li>Assumes that suppliers passed risk down to implementing partners as necessary, to ensure results were achieved</li> <li>The capacity of the suppliers (and implementing partners) and their degree of access to unrestricted funding affected their ability to implement adaptive programming and results-oriented problem-solving</li> </ul>	<ul style="list-style-type: none"> <li>Assumes that suppliers have sufficient control over the WASH results chain to be able to influence results... (was not the case for CLTS which relies on <i>voluntary</i> actions by communities, so actions were largely out of control of suppliers)</li> <li>Assumes that suppliers have sufficient control over their implementing partners to be able to influence results... (was not the case for local government partners which cannot be forced act more quickly = no lever)</li> </ul>	<ul style="list-style-type: none"> <li>Continued learning and best practice in the WASH sector can improve adaptive programming and results-oriented problem-solving over time</li> </ul>
H2a. Introduction of a PbR modality without a requirement relating to inclusivity nevertheless led to the suppliers to addressing the prerequisites for inclusive outcomes	2.2	<ul style="list-style-type: none"> <li>Assumes that the suppliers' dedication to inclusivity ensured they prioritised improving service levels of the poor/vulnerable, despite potential PbR pressures to go for 'low-hanging fruit'</li> </ul>	<ul style="list-style-type: none"> <li>The national contexts can help or hinder the achievement of inclusivity, e.g. Security/ accessibility concerns may prevent suppliers from operating in the most poor/vulnerable areas.</li> </ul>	<ul style="list-style-type: none"> <li>Other national WASH programmes can boost WASH services more broadly</li> </ul>
H3a. The introduction of a PbR modality positively affected the quality of WASH programme outputs and outcomes	2.6	<ul style="list-style-type: none"> <li>Assumes that the average quality of WASH implementation was not always as high as it could be, and that improvement above this baseline was possible and desirable</li> </ul>	<ul style="list-style-type: none"> <li>The national contexts can help or hinder the quality of WASH outcomes, e.g.: <ul style="list-style-type: none"> <li>political momentum behind achieving WASH agenda helps improve the quality of WASH services (e.g. high in</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Continued learning and best practice in the WASH sector can improve the</li> </ul>

		<ul style="list-style-type: none"> <li>Assumes that the scrutiny of the monitoring data which took place as part of the verification process improved the quality of the results achieved</li> <li>Assumes that the additional time spent reporting on results did not detract from the quality of implementation</li> </ul>	<p>Zambia/ Ethiopia – low in Mozambique/Ghana)</p> <ul style="list-style-type: none"> <li>assumes that political/environmental shocks do not hinder operations in country (as they did with Ebola in Liberia, earthquake in Nepal, and insecurity in South Sudan)</li> </ul>	quality of WASH services over time
H4a. The outputs and outcomes implemented were appropriate for achieving the prerequisites for long-term sustainability	2.3, 5.1, 5.2, 5.3	<ul style="list-style-type: none"> <li>Assumes that retaining payments contingent on outcomes being achieved encouraged investment into prerequisites for long-term sustainability <i>beyond programme end</i></li> <li>Assumes that suppliers would (based on their experience) set up the suitable governance arrangements needed for sustainability, even though these were <u>not</u> paid for as an explicit result under PbR (except for SNV)</li> <li>Assumes that suppliers would (based on their experience) choose suitable implementing partners, and build their capacity as needed, to allow these to continue/maintain WASH services beyond programme end</li> <li>Assumes that the need to maintain control over results – and choose implementing partners to maintain such control – could jeopardise the previous assumption</li> <li>Assumes that the pressure to reach output and outcomes by 2018 did not distract efforts from investing in long-term sustainability</li> <li>Assumes introducing a PbR modality without a requirement to link payments to sustainability did not significantly adversely affect the long-term sustainability of outcomes</li> </ul>	<ul style="list-style-type: none"> <li>The national contexts can help or hinder the long-term sustainability of WASH outcomes, e.g.: <ul style="list-style-type: none"> <li>political momentum behind achieving WASH agenda helps ensure the long-term sustainability of WASH services in country (e.g. high in Zambia/ Ethiopia – low in Mozambique/Ghana)</li> <li>assumes that governments / local implementing partners have the capacity to continue/maintain WASH services beyond programme end</li> <li>assumes that political/environmental shocks do not endanger the long-term sustainability of WASH services in country (as they did with Ebola in Liberia, earthquake in Nepal, and insecurity in South Sudan)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Other national WASH programmes can boost WASH services more broadly</li> <li>The national context can shift to increased government and community support for WASH independent of this programme</li> </ul>

### D.3 Evidence within the contribution analysis framework

For each hypothesis, the evaluation team reviewed the extent to which (and how) each were met, and the role of each supplier sub-programme within its wider context – considering how external factors (both enablers and barriers) interacted with the WASH Results Programme activities, as well as any alternative explanations for the results achieved.

Table 9 details the underlying evaluation evidence for each hypothesis, which supports the overall assessment of the WASH Results Programme presented in Section 4 of the main body of the report.

**Table 9: Evidence for each hypothesis**

Outcome hypothesis	Evidence of assumption being realised (programme contribution)	Alternative routes to impact (role of external factors)
<b>H1. The introduction of a PbR modality helped achieve intended outputs and outcomes</b>	<i>[See conclusions under H1.1 / H1.2 / H1.3 below]</i>	
<b>H1.1. The programme and its PbR modality allowed flexibility of implementation approach within the sub-programmes, which helped achieve output and outcome objectives</b>	<p>Consistent evidence of increased flexibility at management level helped, or at least did not hinder, in achieving objectives, with the exception of baseline restrictions. However, no improvement or a worsening of flexibility was reported at local implementation level.</p> <p>The degree of increased flexibility was highly dependent on context: for all suppliers the tight timetable for the output phase somewhat limited options; for the SWIFT consortium management limited flexibility.</p> <p>The use of this flexibility took different forms, including the ability to add or drop programme elements and to move budget between line items. Where flexibility was not possible due to the way targets were constructed under PbR, including the need for appropriate baselines, this could hinder achievement.</p>	External factors did not appear to play a significant role.
<b>H1.2. Stronger monitoring systems as a</b>	Near unanimous feedback that the M&V contributed to supplier monitoring systems being strengthened, and data becoming more credible. Inconsistent evidence on whether	No information on whether a push towards stronger monitoring in the WASH sector affected supplier systems.

<b>result of the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	more credible data led to changes in programming improvements, with some positive examples from SSH4A in Tanzania and for Practical Action's work under SWIFT in Kenya. Also, the tight timetable in the output phase limited possibilities to make adjustments.	
<b>H1.3. The results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	Consistent evidence that the accountability for results focused the minds of all supplier staff on results. Several cases where features of the PbR modality contributed to partners adapting their approach to improve effectiveness. SSH4A reported more instances of this than SWIFT or SAWRP, though all suppliers engaged in adaptive programming to some degree.	The fact that adaptive programming is considered best practice in the sector contributed to this finding.
<b>H2. Introduction of a PbR modality without a requirement relating to inclusivity nevertheless led the suppliers to address the prerequisites for inclusive outcomes.</b>	It appears that PbR neither helped nor hindered a focus on inclusion, though for SWIFT and SAWRP the effect could only be assessed in terms of programme design (due to lack of data on the final distribution of programme benefits).	External factors did not appear to play a significant role. All suppliers had a pre-existing focus on inclusivity in their programme designs.
<b>H3. The introduction of a PbR modality positively affected the quality of WASH programme outputs and outcomes</b>	Inconsistent evidence between and within suppliers, with evidence found of PbR strengthening incentives for quality outputs in some countries but not in others.	No information on whether a push towards continued learning and best practice in the WASH sector affected programming quality.
<b>H4. The outputs and outcomes implemented were appropriate for achieving the prerequisites for long-term sustainability</b>	Reasonable evidence of positive design elements encouraging sustainability, as well documentation of sustainability risks. Across suppliers, sustainability risks were generally higher at the institutional level. Many risks to sustainability were not within the suppliers' control.	Funding and staffing shortages for the local governments where the WASH Results Programme was operating increased challenges for the long-term sustainability of programme achievements.

## Annex E Evidence matrices

The tables below present the high-level summarised evidence matrix for each phase of the contribution analysis. Each supplier is presented in turn. Annex F describes in detail how the contribution analysis was undertaken and which data sources were consulted under each phase.

**Table 10: Evidence matrix for SAWRP**

Outcome hypothesis	Phase I (midline and documentation)	Phase II (country case study)	Phase III (other country interviews)	Phase IV (roundtables and report comments)	Strength of evidence
<b>H1. The introduction of a PbR modality helped achieve intended outputs and outcomes</b>	<i>[Same as Phase II]</i>	There was generally quite consistent praise for the modality, and the inclusion of the output phase from staff. In particular, SAWRP staff at country and district levels highlighted stronger accountability for outcomes and a shift in mind-set. That said, staff explicitly highlighted that the implementation did not change greatly. The inclusion of the outcome phase meant they focused more on sustainability, but the type of activities they carried out to ensure sustainability did not represent a great departure from normal programming. Furthermore, all partners in Pakistan worked with implementing partners – and the PbR features/requirements of the contract (with the exception of monitoring) were not cascaded into their grant agreements.	Previous finding broadly corroborated by the Bangladesh country programme manager interviews, with the important note that elements of the PbR modality were cascaded to WaterAid's implementing partners – triggering some changes in the relationship and implementation.	<i>[See conclusions for H1.1 / H1.2 / H1.3 below]</i>	<i>[See H1.1 / H1.2 / H1.3 below]</i>
<b>H1.1. The programme and its PbR modality allowed flexibility of implementation approach within the sub-programmes, which helped achieve output and outcome objectives</b>	<i>[Same as Phase II]</i>	Budget flexibility and the ability to change activities at the country programme management level were cited as benefits – though this benefit was not cascaded to the implementing partners.	<i>[Unchanged from Phase II]</i>	Consistent evidence of increased flexibility at management level, though limited data from implementation level actors in Bangladesh.	<b>Strong</b>

<b>H1.2. Stronger monitoring systems as a result of the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	<i>[Same as Phase II]</i>	Near unanimous feedback that monitoring systems were strengthened, but less clear whether the more credible data led to changes in programming improvements on the ground.	<i>[Unchanged from Phase II]</i>	Consistent evidence that verification strengthened monitoring.	<b>Strong</b>
<b>H1.3. The results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	<i>[Same as Phase II]</i>	Limited evidence that there were changes made in implementation by the implementing partners in response to new challenges beyond what would be expected in normal programming.	<i>[Unchanged from Phase II]</i>	Consistent feedback from country programme level that there were limited changes in programming made beyond what would be expected in normal programming. Limited insight from field-level staff.	<b>Suggestive/moderate</b>
<b>H2. Introduction of a PbR modality without a requirement relating to inclusivity nevertheless led the suppliers to address the prerequisites for inclusive outcomes</b>	Some evidence that there was an inclusion focus beyond what was contractually required.	<i>[Unchanged from Phase I]</i>	Partners more explicitly targeted marginalised areas, and there was an explicit subsidy component in the WaterAid programme.	Consistent interview evidence on how the PbR modality affected the equity focus, but limited quantitative evidence to accurately describe the intra-community or intra-household distribution of benefits. Reasonable insight into targeting and programme design elements. PbR.	<b>Suggestive/moderate</b>
<b>H3. The introduction of a PbR modality positively affected the quality of WASH programme outputs and outcomes</b>	<i>[Same as Phase II]</i>	Some evidence that partners were responsive to quality elements embedded in the payment indicators – and that this led to changes in implementation. Similarly, such an explicit focus on outcomes was seen to trigger changes.	<i>[Unchanged from Phase II]</i>	Some evidence that PbR positively affected quality, though there was some disagreement among stakeholders.	<b>Suggestive</b>
<b>H4. The outputs and outcomes implemented were appropriate for achieving the prerequisites for long-term sustainability</b>	Some evidence that activities were appropriate for achieving sustainability prerequisites – though clear risks also highlighted.	Sustainability risks were present – though mostly associated with local government capacity (something that was arguably beyond the control of suppliers).	<i>[No additional insight from Bangladesh stakeholders]</i>	Consistent evidence of positive design elements, as well as clear sustainability risks, based on insights from Pakistan. Limited insights for Bangladesh.	<b>Moderate</b>



**Table 11: Evidence matrix for SSH4A**

Outcome hypothesis	Phase I (midline and documentation)	Phase II (country case study)	Phase III (other country interviews)	Phase IV (roundtables and report comments)	Strength of evidence
<b>H1. The introduction of a PbR modality helped achieve intended outputs and outcomes</b>	<i>[See conclusions in H1.1 / H1.2 / H1.3 below]</i>				
<b>H1.1. The programme and its PbR modality allowed flexibility of implementation approach within the sub-programmes, which helped achieve output and outcome objectives</b>	Flexibility in budget allocation and choice of location cited as benefits (midline)	More flexible budgets cited as benefits at country manager level (Uganda and Tanzania) but less flexibility for LCB level in Uganda	SSH4A HQ in Nairobi cited flexibility in budget allocation. SNV country teams felt that PbR had provided them with more flexibility. Mixed picture from LCBs / government.	More flexible budgets cited as benefits at country manager level (across countries) but less flexibility for local implementation level (insight only for Uganda).	<b>Moderate</b> , because insights gained from many countries, but unable to triangulate with a perspective other than SNV.
<b>H1.2. Stronger monitoring systems as a result of the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	Clear evidence from MV reports of surveys being strengthened, but less clear whether the more credible data led to changes in programming improvements on the ground.	Surveys have been strengthened in Uganda and Tanzania. Several examples of cases where having better data has led to changes in programming (see H1.3).	Surveys have been strengthened across countries. SNV has seen benefits of surveys compared to routine monitoring.	Clear evidence from MV reports of surveys being strengthened, and several examples where having better data has led to changes in programming (see H1.3).	<b>Strong</b> , because insights from many countries was triangulated with MV findings.
<b>H1.3. The results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	<i>[No information from this phase]</i>	Several examples of adaptive programming but it does not appear different from BAU.	Several examples of adaptive programming, and in some cases this differed from BAU.	Several examples of adaptive programming, and in some cases this differed from BAU.	<b>Moderate</b> , because insights gained from many countries, but unable to triangulate with a perspective other than SNV.
<b>H2. Introduction of a PbR modality without a requirement relating to inclusivity nevertheless led the suppliers to address the prerequisites for inclusive outcomes</b>	<i>[No information from this phase]</i>	It appears that PbR neither helped nor hindered a focus on inclusion.	Corroborating evidence that PbR neither helped nor hindered a focus on inclusion.	It appears that PbR neither helped nor hindered a focus on inclusion.	<b>Suggestive</b> , because insights only gained for a few countries.
<b>H3. The introduction of a PbR modality positively affected the quality of WASH programme outputs and outcomes</b>	<i>[No information from this phase]</i>	Evidence from Tanzania that PbR strengthened incentives for quality outputs through a focus on	Mixed evidence that PbR has strengthened incentives for quality	Mixed evidence that PbR has strengthened incentives for quality	<b>Suggestive</b> , because insights only gained for a few countries.

		(and payment for) the sustained use of latrines and handwashing facilities. No information for Uganda.	outputs in some countries but not in others.	outputs in some countries but not in others.	
<b>H4. The outputs and outcomes implemented were appropriate for achieving the prerequisites for long-term sustainability</b>	<i>[No information from this phase]</i>	Sustainability prospects are good at community level – but there are risks at institutional level.	Risks to sustainability at institutional level flagged across countries.	Sustainability prospects are good at community level, but there are risks at institutional level.	<b>Strong</b> , because insights from Uganda and Tanzania were triangulated with sustainability indicator results from all countries.

**Table 12: Evidence matrix for SWIFT**

Outcome hypothesis	Phase I (midline and documentation)	Phase II (country case study)	Phase III (other country interviews)	Phase IV (roundtables and report comments)	Strength of evidence
<b>H1. The introduction of a PbR modality helped achieve intended outputs and outcomes</b>	<i>[See conclusions in H1.1 / H1.2 / H1.3 below]</i>				
<b>H1.1. The programme and its PbR modality allowed flexibility of implementation approach within the sub-programmes, which helped achieve output and outcome objectives</b>	A clear evidence trail of DFID showing 'flexibility' in taking on risk <i>post-hoc</i> , but no strong link seen to programming.	Because the different partner programmes were so varied some cited great flexibility, while others felt no change. Some cited that the inflexibility of the targets meant they could not respond to changes in context.	In DRC budget flexibility was an important benefit, though this was not cascaded to lower implementation levels and was tempered by other programme constraints – mainly time.	Mixed picture, with operational flexibility cited as a benefit by some SWIFT partners, but not others.	<b>Moderate/strong</b>
<b>H1.2. Stronger monitoring systems as a result of the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	Clear evidence from MV reports of issues with monitoring systems being identified and resolved.	Unanimous feedback that monitoring systems were strengthened. Less clear whether the more credible data led to changes in programming improvements on the ground.	Clear evidence noting the MV's role in strengthened monitoring. Link to programming less clear.	Consistent feedback that verification strengthened monitoring, but no evidence that this led to changes in programming.	<b>Strong</b>
<b>H1.3. The results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes</b>	<i>[No information from this phase]</i>	Mixed picture – some isolated examples (mainly Practical Action) of cases where suppliers cited adaptive programming,	Mixed picture – isolated examples of adaptive programming, noting that other constraints (time and budget) restricted their	Consistent feedback from country programme level that there were limited changes in programming made beyond what would be expected in normal programming, except for	<b>Moderate</b>

		though not especially widespread.	ability to 'do' adaptive programming.	some positive examples for Practical Action. Limited insight from field-level staff.	
<b>H2. Introduction of a PbR modality without a requirement relating to inclusivity nevertheless led the suppliers to address the prerequisites for inclusive outcomes</b>	<i>[No information from this phase]</i>	Limited insight on whether PbR helped or hindered inclusivity – though some cases where positive design elements were highlighted. No strong quantitative evidence.	<i>[Limited additional insight]</i>	Consistent interview evidence on how the PbR modality affected the equity focus. Limited quantitative evidence on intra-community or intra-household distribution of benefits, but some insight related to targeting and programme design elements. PbR.	
<b>H3. The introduction of a PbR modality positively affected the quality of WASH programme outputs and outcomes</b>	<i>[No information from this phase]</i>	Limited examples of PbR positively affecting quality from the isolated cases visited.	Very limited indication that the indicators and the verification incentivised an additional focus on quality	Reasonable confidence that PbR positively affected quality, based on isolated cases visited in Kenya (though only a small portion of infrastructure was reviewed). Very limited insights from DRC.	<b>Suggestive/moderate</b>
<b>H4. The outputs and outcomes implemented were appropriate for achieving the prerequisites for long-term sustainability</b>	The light-touch sustainability assessment raised some clear sustainability concerns.	Mixed picture – with some clear and present sustainability risks.	<i>[Limited additional insight]</i>	Reasonable evidence of positive design elements, as well as clear sustainability risks, based on insights from Kenya, though the evidence for DRC is more limited.	<b>Suggestive</b>

## Annex F Overview of data collection

### F.1 List of supplier documents consulted

Documents consulted during the endline evaluation	
SSH4A	SSH4A Midline Evaluation Richard's midline report
	SSH4A Midline Evaluation Khasifa's supplementary study on LCBs
	SSH4A Supplier documentation Endline survey reports (x8 for each country)
	Itad MV country visit reports
	SSH4A Verification reports for RP10 and RP11
	SSH4A Payment decision meeting notes RP5
	SSH4A Payment decision meeting notes RP6 (dec2015)
	SSH4A Payment decision meeting notes RP10
	SSH4A Payment decision meeting notes RP11
	Sustainability indicator reports (2017)
	Sustainability indicator reports (2018)
SAWRP	SWARP: Monitoring, Verification and Evaluation Framework for Project Outcomes (MVOC)
	SWARP: Child handwashing mid and endline survey report
	SWARP: Minutes of the Child handwashing mid and endline payment decision meeting
	SWARP: Outcome survey 1 verification report
	SWARP: Household midline survey report
	SWARP: Outcome survey 1 and HH midline payment decision minutes
	SWARP: Outcome survey 2 verification report
	SWARP: Outcome survey 2 payment decision
	SWARP: Internal 6 month narrative report (July-Dec 2016)
	SWARP: Pakistan First Round SAF, Consolidated Report
SWIFT	SWIFT Verification report Q1 2016
	SWIFT Payment decision meeting Q1 2016
	SWIFT Verification report Q4 2016
	SWIFT Payment decision meeting Q4 2016
	SWIFT Verification report Q2 2017
	SWIFT Payment decision meeting Q2 2017
	SWIFT Verification report Q4 2017
	SWIFT Payment decision meeting Q4 2017
	SWIFT Consortium Quarterly Technical Report Q1 2016
	SWIFT Consortium Progress Report Q1&Q2 2017

## F.2 List of people interviewed

**Table 13: People interviewed during SAWRP Pakistan case study**

Organisation	Position
Plan International	CCU Manager
	Plan acting Country Director
	Plan programme manager
	Plan M&E
	Plan M&E
	Finance
	WASH specialist
WaterAid	WaterAid Country Director
	Programme Manager
	M&E
	BCC manager for SAWRP
LPP	Project manager
	Director of implementation
Muslim Aid	National programme manager
	WASH programme manager
LGCD Punjab	Provincial coordinator of WASH - Punjab
LGCD KPK	KPK assistant provincial WASH coordinator
LGCD Sindh	Deputy director - local government - Sindh
Lead verifier	Lead verifier for SAWRP
Country verifier	Country verifier for SAWRP
AWARE	programme head
	District coordinator
NRSP – Sindh	Project managers
	M & E
NRSP	NRSP national M&E officer

**Table 14: People interviewed during SSH4A Uganda case study**

Organisation	Title / affiliation
SNV Uganda	SSH4A Country Manager
	Programme Adviser West Nile
	Programme Adviser Rwenzori
Country Verifiers	Country Verifier Uganda
LCB - EMESCO	NGO manager
Local government – Kibaale District	CAO
	District planner
	Health Inspector
Local government – Mubende District	DHI
	Deputy CAO
	ADHO
LCB - Rural Health Care Foundation (Mubende)	Health Assistant

LCB - Rural Health Care Foundation (Mubende)	Health Assistant
Local government – Kenjojo District	DHI - Kyegegwa
LCB - HEWASA	LCB partner
Local government – Kyegegwa District	CDO - Officer in charge of sanitation
Local government – Kyegegwa District	District Subc HA
LCB - URDT (Kyegegwa)	Project officer
UNICEF	WASH Chief, Uganda
USF	Project manager

**Table 15: People interviewed during SSH4A Tanzania case study**

Organisation	Title / affiliation
SNV Tanzania	SSH4A Country Manager
	Programme Adviser Arusha
Local government – Babati District	District-level government officials
	Ward-level government officials
	Sub-village government officials
	CLTS facilitators
Local government – Karatu District	District-level government officials
	Ward-level government officials
	Sub-village government officials
	CLTS facilitators

**Table 16: People interviewed during SWIFT Kenya case study**

Organisation	Title / affiliation
Oxfam	Oxfam WASH team leader Lodwar
	OXFAM Sanitation technician Lodwar
	WASH manager OXFAM
	SWIFT Global Project Manager
	MEAL officer OXFAM (Former)
Government – Sub county level	PHO Kakuma ward Turkana West at sub-county public health office
	PHO Leta ward Turkana West at sub-county public health office
	PHO Lutea ward Turkana West)
KAWASEPRO	
	Secretary General
	Technician
	Chairman
Sub-county water office	Sub county water officer
Supplier – community level (CHV)	
Practical Action	Programme management
Practical Action	Programme management
Practical Action	M&E



Practical Action Turkana	WASH technician
Lorghum community	Chief Lorghum
Lorghum community	Chairman Lorghum WUA
Lorghum community	Secretary Lorghum WUA
Turkwel community	Senior chief Turkwel
Turkwel community	Water controller Turkwel water association
Turkwel community	Secretary Turkwel water association
Turkwel community	Plumber Turkwel water association
Lolupe community	CHEW Lolupe
Lolupe community	CHEW Lolupe
Lolupe community	CHEW Lolupe
Lolupe community	CHV Lolupe
Lolupe community	CHEW Lolupe
Lolupe community	CHEW Lolupe
Sanergy	business development/ grant management
WSUP	M&E officer
WSUP	Project manager
WSUP	Project officer
Nairobi Water Company	Community Development Manager
Nairobi Water Company	Nairobi Water staff at Dandora site
WSUP field visit to Dandora	WSU Dandora site office promoter
Country verifier	
Country verifier	
BBCMA	

**Table 17: 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	Guy Howard	Senior Responsible Owner (SRO)
	Stephen Lindley-Jones	Senior WASH officer
ITAD (MV)	Kathi Welle	Deputy-lead of MV Team
	Rachel Norman	Deputy-lead of MV Team (maternity cover)
	Andrew Robinson	LV for SSH4A
IWEL (MV)	Don Brown	Lead of MV Team
	Alison Barrett	LV for SAWRP
	Joe Gomme	LV for SWIFT
Plan (SAWRP)	Teia Rogers	Lead for SAWRP
	John Dean	M&E officer
WaterAid (SAWRP)	Siddiq Ahmed Khan	WaterAid country director
ODI (SWIFT)	Nat Mason	SWIFT learning lead
	Ian Langdown	Research officer at ODI

WEDC (SAWRP)	Susie Goodall	SAWRP learning officer
	Mimi Coultas	SWIFT learning lead
SNV (SSH4A)	Antoinette Kome	Lead for SSH4A (2014)
	Anne Mutta	Lead for SSH4A (2015-2018)
Oxfam (SWIFT)	Joanna Trevor	Lead for SWIFT (2016-2018) DRC country manager (2014-2016)
Tearfund (SWIFT)	Rachel Stevens	Tear fund UK manager

**Table 18: Country managers interviewed remotely**

Organisation	Country manager position	Name
SNV (SSH4A)	Ethiopia	Andualem Anteneh
	Kenya	Fanuel Nyaboro
	Mozambique	Nelson
	Nepal	Ratan Budhathoki
	Tanzania	Jackson Wandera
	Zambia	Solomon Mbewe
Oxfam (SWIFT)	DRC	David Marques (ex-SWIFT manager)
Tearfund (SWIFT)	DRC	Hebdavi Muhindo
Plan (SAWRP)	Bangladesh	Fadia Sultana (SAWRP manager)
	Bangladesh	Zillur Rahman (head of WASH)
WaterAid (SAWRP)	Bangladesh	Khairul Islam (country director)

## Annex G Summary of verification processes

This annex describes the design of the verification procedures for each of the suppliers and relates particularly to the ‘relevance’ aspects of the evaluation.

### G.1 Verification framework

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 TOR but was chosen by the MV provider to match the budgetary envelope of the TOR. 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 to 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 was conducted ahead of the first full verification cycle to map which internal monitoring and reporting systems would generate the evidence needed for verification. If systems were deemed to be insufficient, corrective action was recommended by the verifiers. The systems appraisal was repeated ahead of each verification cycle until the systems were deemed to meet the required standard for evidence generation, and/or if evidence requirements changed over time.
- **Desk-based verification of supplier-generated evidence:** First, a list of evidence requirements was drawn up by the MV team (the ‘Form 2’), tailored to each verification round as necessary. This evidence was 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 carried 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 took place in parallel and informed the conclusion by the verification team:

- **An initial verification report** per verification round reported the results of the verification activities.
- **This was discussed in payment decision meetings** involving DFID, the supplier leads, and the MVE team, where either payment decisions were made or further evidence was requested by DFID to inform a payment decision.
- **An After-Action Review** was frequently held thereafter to identify lessons and agree on actions to take in forthcoming verification rounds.

**Table 19: Duration of verification cycle for each supplier during the output phase**

SWIFT	SAWRP	SSH4A
<ul style="list-style-type: none"> <li>Several weeks for SWIFT to provide evidence</li> <li>From evidence submission: two weeks for clarifications on evidence</li> <li>2.5 weeks (in parallel with above) for country report submission</li> <li>1.5 weeks for supplier report submission and quality assurance</li> <li>Verification report was submitted one week before payment meeting</li> </ul>	<ul style="list-style-type: none"> <li>One week for SAWRP to provide the draft database for sampling</li> <li>Five working days for provision of sample secondary evidence to country verifiers</li> <li>2.5 weeks (approx.) to verify evidence, ask for clarifications, and write verification reports</li> <li>Verification report was submitted 10 days before the payment meeting</li> </ul>	<ul style="list-style-type: none"> <li>Two weeks for field spot checks (when required)</li> <li>Two weeks for the county verification reports</li> <li>One week for the lead verifier to draft the verification report</li> <li>Verification report was submitted one week before payment meeting</li> </ul>
<b>Five- to six-week cycle</b>	<b>Five-week cycle</b>	<b>Three- or five-week cycle</b>

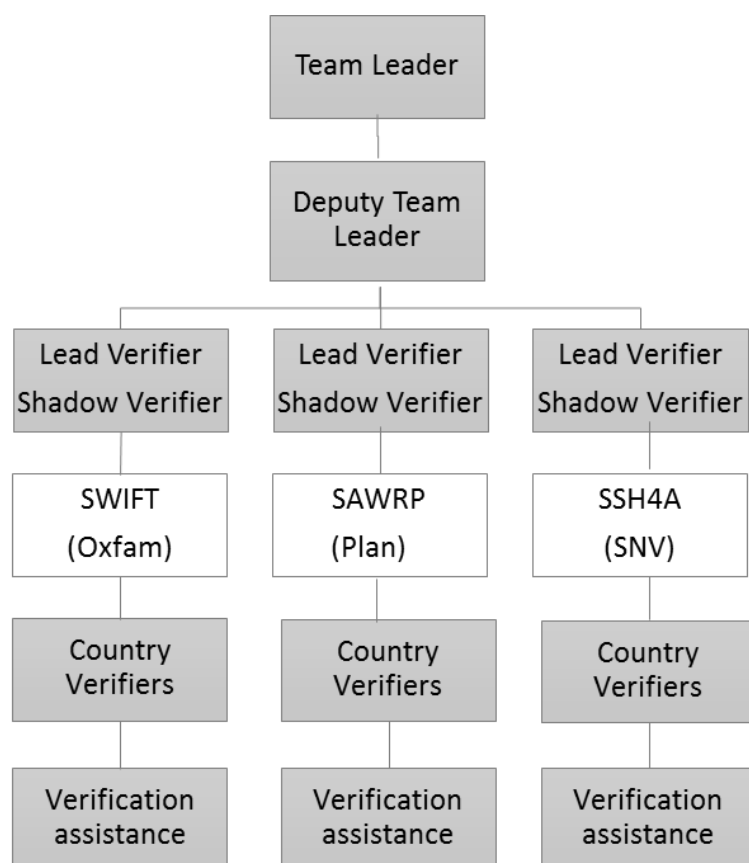
This process was repeated almost every quarter under the output phase, and has been repeated every second or fourth quarter during the outcome phase to date, depending on the supplier (see table below for the actual verification rounds conducted by supplier).

**Table 20: Timing of verification rounds for each supplier**

	2014			2015				2016				2017				2018	
	Output-related payments							Outcome-related payments									
Calendar quarter->	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	
SAWRP		X	X	X	X	X	X				X	X	X		X		
SSH4A		X X	X		X	X	X				X		X		X	X	
SWIFT	X	X	X	X	X	X	X	X			X		X		X		

In addition to this verification function, the verification team also had **an advisory function**: they provided 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 aimed to flag up quality issues, and to assess where there was room for improvement for the supplier in regard to delivering WASH services. This advice did not directly affect payment but if key risks were identified, action plans were agreed with suppliers through After-Action Reviews that set out mitigation measures.

The verification team comprised one lead verifier, one deputy verifier (previously called the ‘shadow verifier’) and a series of national country verifiers for each supplier (see Figure 4). The relative level of effort required of different members of the verification team was tailored to the tasks involved in each verification round.

**Figure 4: Organogram of the MV team**

### 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 procedures where possible, and appraising these allowed verifiers to check the reliability of reported data. Internal quality assurance 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.

### 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) were built around the same elements: for each verification indicator, the form specified the indicator definition, data source, and data requirements for suppliers, and the methodology of analysis for the verifier. Indicators usually included 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 included the

methodology for how a payment decision was made, based on the aggregate analysis of all the indicators which pertained to that deliverable, such as whether payment was proportional to the number of results achieved or whether it was 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. As each supplier was assigned a different lead verifier, 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.

## **Tailoring systems appraisals to suppliers**

Systems appraisals were intended to check whether internal monitoring and reporting systems were 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:

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 Reviews. 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. 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 posed risks for the reliability of the outcome surveys based on these baseline findings.

## **Country verifier field visits**

Originally, the systems-based approach to verification was only designed around a 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 quality assurance 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.

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.

### **Consequences of ‘failing’ a verification round**

While the verification process was designed as a step-by-step process, all three lead verifiers 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. Lead verifiers felt that without this flexibility, far fewer results packages would have passed each verification cycle.

## Annex H Stakeholder comments on the report

Suppliers were provided with the opportunity to comment on the Case Study reports pertaining to their consortium, and two opportunities at stages in the report drafting. In the vast majority of cases the suppliers comments were resolved. However there were some comments where the evaluation team's appraisal differed from the Supplier comments substantially. These are mapped below including the evaluation team's reply to SAWRP on these points. SAWRP were invited to submit further comments in reply which would have been included in this Annex.

Section	Text commented on	Comment from SAWRP	Reply to SAWRP
4.2.1.3	"Though not reflected in the results data, suppliers in Pakistan reported an ODF conversion rate of around 98% (note that this has not been verified by e-Pact). By global standards, this is exceptional and is therefore controversial."	SAWRP: "Comment in previous version not addressed - "controversial" suggests negative but not in scope (wording could instead be e.g. level of achievement would be exceptional)"	The evaluation team remain very sceptical about the ODF claims. While we do not doubt that the communities were verified ODF by government, and that Plan staff followed the steps in this regard. We do have concerns around how meaningful the ODF claim is related to: i) the definitions and standards used by local government; and the fact these vary across and even within districts, ii) in some cases there was 'bulk' ODF verification – i.e. communities would be bulk verified based on a small sample. In the case of 1 local partner in Pakistan 150 communities were verified on the basis of the actual verification of ~10 communities, and iii) the RCT qualitative results from ODF communities raised serious concerns about the standards applied and found evidence of slippage.
4.4.2. on table 18 points 4 and 7	Ratings of 'somewhat' and 'unlikely'	SAWRP: "Still somewhat? See previous version - mothers/women (Ideal Mother Groups) were primary target group of our BCC approach during the outcome phase in particular." And "See previous version - This seems rather unrepresentative, when elsewhere it states that 98% of SAWRP villages achieved ODF villages, which by its nature is very inclusive and includes working with Government, despite not being a PbR target. SAWRP partners worked with community organisations and local government on inclusion issues, although this was often challenging, as the Local Government system is not yet fully functional in Pakistan, since roll-out of the new system began in 2013. These ratings should also take account of how receptive local government was to policy change."	The comments given previously were considered by the case study team who completed the framework (Jeremy Colin and Zach White). Ultimately the ratings were kept as they were.  On these points we invite Plan to prepare a note/ comment that would be included verbatim in an Annex and clearly referenced at the appropriate point in the main body of the report.
4.5 on table 19 on DEQ 5.3	"In contrast, the rushed output phase in some cases was likely detrimental to efforts to enhance the likelihood of sustainable services."	SAWRP: "Considering the high outcome results is this accurate?"	Yes, in the view of the evaluation team this is a reasonable conclusion across the whole programme. Functionality, as explicitly outlined in the report, is not seen to be the same thing as sustainability. Furthermore, functionality was not high in all places or countries. This statement also draws heavily on the midline findings where there was strong criticism of the rushed nature of the output phase across suppliers and at many different implementation levels. We stand by the statement as a fair summative assessment across the WRP.

## Annex I VFM analysis based on price

At midline the evaluation team faced significant challenges in obtaining the data necessary to conduct a VFM analysis in line with the approach proposed at design stage. The VFM analysis framework proposed at design stage built on the approach developed under the DFID 'VFM WASH project' – see Section 5.4.2 of the Design Document for an overview of the approach

The VFM framework proposed at design stage distinguished between a 'light-touch' VFM framework based on price data (see Section 5.4.3 of the Evaluation Design Document ) and a more 'in-depth' VFM analysis which would use actual cost data (see Section 5.4.4 of the Evaluation Design Document). At design stage there was caution regarding using the light-touch approach as it was seen as highly likely that prices and costs for certain areas would differ substantially. Similarly, concerns were raised about the feasibility of the VFM analysis given the commercial nature of the supplier contracts, and the fact that it was seen as unlikely that suppliers would be willing to share the data necessary to conduct the analysis. For example, SAWRP explicitly requested the following contract clause, which was agreed with DFID (emphasis added):

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

The data requirements for the 'in-depth' VFM analysis, alongside a feasibility assessment of the VFM analysis, are detailed in Annex C.6 of the Evaluation Design Document.

These concerns were realised at the midline assessment where the suppliers were unwilling to share detailed cost data and furthermore highlighted that not all of the cost data collected would be suitable for the purposes of the 'in-depth' VFM analysis. As such, ahead of the endline data collection it was agreed with DFID that some of the evaluation questions (DEQs 3.2 and 3.3.) relating to VFM would be removed while others had their wording updated (DEQs 3.4. and 3.6.) to focus on the drivers of VFM rather than aiming to produce VFM metrics. It was also agreed at this stage that a 'light-touch' VFM analysis would be conducted on the price data available.

**The 'light-touch' VFM analysis was conducted and included in a draft report prepared for DFID. The analysis was conducted as it was part of the agreed methodology. Though the evaluation team has serious misgivings about how meaningful an analysis based on price data is and the results of the analysis were not included in subsequent drafts as there were concerns that if published the results would likely be misleading should they be quoted out of context.**

## Annex J Annotated bibliography on the application of PbR in WASH development (2013-2018)

This literature review covers literature produced in the last 5 years within the field of payment by results (PbR), predominantly in the field of WASH development aid. Starting in 2013, DFID published the business case for the WASH Results programme<sup>46</sup> one year prior to DFID's official strategy was published to sharpen incentives to perform through PbR<sup>47</sup>. The business case recognises that a WASH development programme of this size based on a PbR mechanism would be a risk for both DFID and potential suppliers, inexperienced in its management and implementation, but would provide a valuable opportunity to generate evidence on the use of PbR in the WASH development sector. At the time, DFID anticipated the additional costs of monitoring and evaluation; the disadvantage of excluding smaller suppliers without an appetite for greater financial risk; and the challenge of negotiating PbR contracts that address country-specific risks, for e.g. risks beyond the control of the supplier in fragile contexts. On the other hand, if implemented right, a PbR WASH programme offered the opportunity to unload greater risk onto the supplier; create positive incentives to deliver results; encourage innovation in the delivery of result; strengthen accountability and place a focus on performance and efficiency. By then end of 2013, 71% of DFID centrally issued contracts from DFID had a PbR component. In 2014, DFID declared their ambitions to push innovation in the way that PbR contracts are designed to shift towards longer-term results, output-based payments. At this time, DFID was only beginning to explore outcome-based aid to address the sustainability of output, mainly in the fields of health, infrastructure and education. DFID's 2014 strategy laid out clear objectives to remain at the forefront of PbR financing; expand the evidence base for how PbR can deliver better results in different sectors and develop internal capacity to manage and guide rigorous, independent and comparable evaluations for what works best in PbR financing. It is clear from a review of recent DFID publications on PbR that PbR is the preferred payment modality and that DFID has followed through on their commitment to expand the evidence base for PbR in different sectors.<sup>48</sup>

Monitoring, evaluation and verification (MVE) of DFID WASH Results programme still stands to offer a great deal of evidence on how to do PbR right in the WASH sector, as, even today, most of the evidence on the effectiveness of PbR in delivering better, more cost-effective results still comes from the fields of health and education. There are signs that PbR is gaining traction amongst other key donors in the WASH sector, for example USAID<sup>49,50</sup> and Australian Aid<sup>51</sup> funded or are currently funding WASH programmes with a PbR mechanism in SE Asia and SSA. The World Bank, through its Program for Results<sup>52</sup>, is currently funding a WASH programme in Vietnam, which includes a PbR element.

DFID still stands out as the largest donor that is pushing PbR as the predominant contracting modality. Within the WASH sector, DFID has funded 3 large WASH programmes with a PbR mechanism, Support to Rural Water Supply, Sanitation & Hygiene in Tanzania (2014-2022). £78.6m, around half of the contract value, is to be paid out to Local Government Authorities based on the

<sup>46</sup> DFID. (2013). Business Case and Intervention Summary: Water, Sanitation and Hygiene (WASH) Results Programme. Available at: <https://devtracker.dfid.gov.uk/projects/GB-1-203572/documents>

<sup>47</sup> DFID. (2014). Sharpening incentives to perform: DFID's Strategy for Payment by results. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/323868/Sharpening\\_incentives\\_to\\_perform\\_DFID's\\_Strategy\\_on\\_Payment\\_by\\_Results.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/323868/Sharpening_incentives_to_perform_DFID's_Strategy_on_Payment_by_Results.pdf)

<sup>48</sup> See for example, Clist, 2017 and Duvendack, 2017.

<sup>49</sup> See for example, Jenkins, M. (2016). How can we reach the poor with market-based approaches? 2016 Colorado WASH Symposium, University of Colorado Boulder. Available at: [https://www.colorado.edu/washsymposium/sites/default/files/attached-files/ucolorado\\_reaching\\_the\\_poorest\\_mar1\\_2016\\_v1\\_to\\_post.pdf](https://www.colorado.edu/washsymposium/sites/default/files/attached-files/ucolorado_reaching_the_poorest_mar1_2016_v1_to_post.pdf)

<sup>50</sup> USAID. (2018). WAS-FIN Kenya project Brief – Access to commercial finance for WASH in Kenya: Looking back on SUWASA and to the future with WASH-FIN. Available at: <https://files.globalwaters.org/water-links-files/WASH-FIN-Kenya-Project-Brief.pdf>

<sup>51</sup> Thrive Networks. Available at: <http://thrivenetworks.org/output-based-aid-delivering-wash-services-vietnam-ensuring-sustainability-reaching-poor/>

<sup>52</sup> WB. Program-for-Results Financing (PforR). Available at: <http://www.worldbank.org/en/programs/program-for-results-financing>

delivery of results, with particular emphasis on maintenance of rural water supply infrastructure. The 2015 Annual review of this programme described a delay in the start of the PbR mechanism due to insufficient baseline data, which makes verification of results impossible. DFID is also funding Accelerating Sanitation and Water for All (ASWA I) in neglected, Off-Track Countries. The business case for the ASWA I programme describes a fairly simple implementation model for the PbR component. DFID will make semi-annual payments. Starting with the second payment, 10% of funds will be payable upon demonstration of results. DFID is also funding Supporting the Transformation of Rural WASH Service Delivery in Mozambique. The PbR component is planned to start 18 months into implementation once performance indicators and incentive structure is finalised with the government of Mozambique.

The key lines of evidence to refer to for how PbR can improve WASH results are – summaries are provided further down:

Clist, P. (2017). Full Report: Review of Payment by Results in DFID: Establishing the Evidence Base: Objective 2: What works for Payment by results Mechanisms in DFID Programs. Available at:  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/684278/full-report-UEA2-merged.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684278/full-report-UEA2-merged.pdf)

Duvendack, M. (2017). Full Report: Review of Payment by Results in DFID: Establishing the Evidence Base: Semi-systematic review to understand Payment-by-Results mechanisms in developing countries. Available at:  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/684277/full-report-UEA1-merged.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684277/full-report-UEA1-merged.pdf)

Castalia. (2015). Review of Results-Based Financing Schemes in WASH. Available at:  
[https://www.ircwash.org/sites/default/files/rbf\\_in\\_wash\\_final\\_report\\_full\\_jan\\_2015.pdf](https://www.ircwash.org/sites/default/files/rbf_in_wash_final_report_full_jan_2015.pdf)

Rodriguez, Diego J., Suardi, M. A., Ham, M., Mimmi, L. M., Goksu, A. J. (2014). Applying results-based financing in water investments (English). Water papers; Water partnership program (WPP). Washington, DC: World Bank Group.  
<http://documents.worldbank.org/curated/en/862681468326999086/Applying-results-based-financing-in-water-investments>

Global Partnership on Output-Based Aid. (2016). Water Sector Experience of Output-Based Aid. World Bank, Washington, DC. Available at:  
<https://openknowledge.worldbank.org/handle/10986/24674>

### J.1.1 DFID WASH projects with PbR

**Table 21 Summary of other DFID-funded WASH Projects**

Name of project	Country of implementation	Period of implementation	Total project funding	Information on PbR disbursement
Support to Rural Water Supply, Sanitation & Hygiene in Tanzania (2014-2022)	Tanzania	2014-2022	£150m	£78.6m to be disbursed under PbR to local governments (focus on maintenance of rural water supply infrastructure)
ASWA I (2013-2019) (can't find the DevTracker page for ASWA II, but it has just started/is yet to start)	Pakistan, Nepal, Bangladesh, Myanmar, Cambodia, Niger, South Sudan, Yemen, Madagascar	2013-2019	£47,252,050	Payments on semi-annual basis. Starting on 2 <sup>nd</sup> payment, 10% disbursed as PbR.
Supporting the Transformation of Rural WASH Service Delivery in	Mozambique	2015-2020	£38m	PbR planned to start 18 months into implementation once performance indicators and incentive structure is finalised with government.

Mozambique (2015-2020)

## Support to Rural Water Supply, Sanitation & Hygiene in Tanzania (2014-2022)

<https://devtracker.dfid.gov.uk/projects/GB-1-204033/documents>

The business case for 'Phase 2: Rural Water Supply and Sanitation Programme, Tanzania' outlines the case for £ 150 million to be spread over the period of April 2014 to March 2019. The division of funding is to be split into 4 areas:

- (1) £75m to support water sector development (construction & rehabilitation of water points and support to national sanitation and hygiene promotion campaigns)
- (2) A PbR modality will be used to disperse up to £66.6m (this changed to £78.6m in implementation) to support Local Government Authorities in the delivery of results, with particular emphasis on maintenance of rural water supply infrastructure.
- (3) £3.4m will be dedicated to monitoring and evaluation, **including the independent monitoring and evaluation of the PbR component**. DFID is also carrying out an evaluation to analyse various points in rural water point maintenance, which is intended to inform the design and implementation of PbR scheme
- (4) £5m will be dedicated to offering technical assistance to key government actors to support institutional strengthening

### How will PbR be implemented?

A phased approach is suggested in the business case. Some Local Government Authorities (LGA's) will have limited capacity and risk not benefiting from the PbR scheme, so PbR will be implemented in a phased manner with these LGA's. In addition the Technical Assistance component will prioritise LGAs with limited capacity and operating under the PbR scheme to optimise the benefits of operating under a PbR scheme.

The PbR mechanism seeks to address the poor accountability and performance of LGAs in delivering and maintaining rural water supply access. It will complement Tanzania's Big Results Now, the GoT's commitment to delivering rural water supply access. Importantly, KPIs have been developed as part of the BRN initiative to improve performance monitoring and accountability in the delivery of rural water supply, considering each actor along the delivery chain. The general theory of change is that collectively PbR and BRN will facilitate conversion of inputs into outcomes.

### What has the 2015 Annual revealed about the PbR mechanism in this WASH programme?

The PbR scheme was delayed due to unavailability of sufficient and accurate data and delays in the recruitment of Technical Assistants for LGAs. A lack of robust baseline data has contributed to the delay in implementing the PbR scheme. Water Point Mapping System and availability of data on water point functionality in real time is insufficient. The PbR scheme, although delayed, has made improving monitoring and data quality a priority.

## ASWA I (2013-2019)

<https://devtracker.dfid.gov.uk/projects/GB-1-203571/documents>

The business case for the Accelerating Sanitation, Hygiene and Water for All in Off-Track Countries planned to support UNICEF with £35.5m over five year period from 2013-2017 with an additional £1.5m earmarked for an independent impact evaluation. The current budget published on DevTracker is £47,252,050 and the expected end date of the programme is 2019. The business case also elaborated plans for DFID to work with UNICEF to strengthen their existing results monitoring systems and include independent verification of results.



The business case for the ASWA I programme describes a fairly simple implementation model for the PbR component. DFID will make semi-annual payments. Starting with the second payment, 10% of funds will be payable upon demonstration of results.

Also see, independent evaluation of ASWA I: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/701093/Accelerating-Sanitation-and-Water-for-All-in-Neglected-Off-track-Countries.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/701093/Accelerating-Sanitation-and-Water-for-All-in-Neglected-Off-track-Countries.pdf)

### **DFID Nigeria business case: Phase 2 of the Sanitation, Hygiene and Water in Nigeria Programme (SHAWN II) – EXPLAINS WHY NOT USING PbR MODALITY**

<https://devtracker.dfid.gov.uk/projects/GB-1-202977/documents>

The business case for this programme cites Winpenny (2013) stating that PbR works best in urban and peri-urban areas and that careful design is required to extend this model to scattered more urban areas. Due to the challenges of operating in rural areas, private operators are far less likely to tender for a PbR-based implementation component of SHAWN. Alternatively, SHAWN awards contracts based on competition using the UN procurement rules.

### **Supporting the Transformation of Rural WASH Service Delivery in Mozambique (2015-2020)**

<https://devtracker.dfid.gov.uk/projects/GB-1-204856/documents>

The business case for this programme describes how the UK will provide £38 million over a 5-year period to improve access to safe water and sanitation in Mozambique, with a particular emphasis on support the GoM's aim to double rural access to safe water by 2019 and increase access to sanitation. Support will take the form of financial aid and technical assistance, where, in part, payment will be based on agreed upon performance indicators, which will be verified by an independent monitoring partner. The PbR modality would only start 18 months into implementation following a kick-off period where financial aid is disbursed as usual. During this kick-off period, DFID will work with the GoM to define performance indicators that incentivize improved services. The most recent Annual Review of the 2016-2017 programme year confirms that the performance indicators were collaboratively developed with national, regional and local levels of government and a monitoring and evaluation framework is in development by the independent verification partners in line with agreed performance indicators.

Also see the last planned independent evaluation report from 2017 for lessons learned. Available on DevTracker.

## **J.1.2 Non-DFID projects with PbR**

**Table 22 Summary of non-DFID-funded WASH projects (2014-2018)**

Name of funder	Name of project	Country of implementation	Period of implementation	Information on PbR disbursement
Australian Aid	Water, Sanitation and Hygiene Output based Aid (WASHOBA)	Lao, Vietnam & Cambodia	2014-2018	Designed to strengthen incentives for household latrine delivery. Service providers payed under OBA modality – per-connection payment for each new last mile client. See CS WASH Fund Projects, n.b. for more details.
USAID	Sustainable Water and Sanitation in Africa (SUWASA)	Kenya	2010-2015	SUWASA assisted Water Service Providers (WSPs) to access funds from KfW and the

				World Bank based on a PbR modality through the Water Sector Trust Fund, an independent body which was also created under the project. Portions of the loans that WSP gained access to were paid off by KfW and The World Bank based on the delivery of agreed-upon results. See USAID, 2018 for further details.
USAID	WASH-FIN	Kenya	2016-2021	This programme will build upon the success of the SUWASA programme. See USAID, 2018 for further details).
GPOBA	Numerous – focus on water supply	Numerous countries in East Asia, the Pacific, South Asia, Sub-Saharan Africa, Middle East, North Africa, Latin America and the Caribbean	Between 2003-2016, GPOBA has supported 12 water supply projects	GPOBA has supported a wide range of service providers (small to large-scale providers, state and private, community and NGO service providers) in varied context under an OBA modality. See GPOBA, 2016 for more details.

CS WASH Fund Projects. (n.d). Water, Sanitation and Hygiene Output based Aid (WASHOBA) – Cambodia, Laos and Vietnam (Australian Aid). Available at:

<http://www.cswashfund.org/our-work/about/grantee-profiles/thrive-networks>

Also see: Thrive's CS WASH Fund project is helping the government to extend water supply networks to households in areas affected by salinity intrusion - Vietnam (Australian Aid). Available at:

<http://www.cswashfund.org/our-work/postcards-from-the-field/oba-action>

Also see other project documents and papers on Thrive Network website. Available at:

<http://thrivenetworks.org/output-based-aid-delivering-wash-services-vietnam-ensuring-sustainability-reaching-poor/>

Also see: Nyugen, H. (2016). The effect of OBA subsidies combined with sanitation marketing (SanMark) on Latrine Uptake Mong Rural Populations in Cambodia. Available at:

<http://iwcan.org/effect-oba-subsidies-combined-sanitation-marketing-sanmark-latrine-uptake-among-rural-populations-cambodia/>

Also see: Civil Society WASH Fund Management Facility/Aguaconsult. (2018). Effective Approaches and Innovations in the Civil Society WASH Fund: Research Report. Available at: [https://www.aguaconsult.co.uk/wp-content/uploads/CSWASHFund\\_2018\\_Effectiveness-WASH-Approaches-Research-Report.pdf](https://www.aguaconsult.co.uk/wp-content/uploads/CSWASHFund_2018_Effectiveness-WASH-Approaches-Research-Report.pdf)

USAID. (2018). WASH-FIN Kenya project Brief – Access to commercial finance for WASH in Kenya: Looking back on SUWASA and to the future with WASH-FIN. Available at:

<https://files.globalwaters.org/water-links-files/WASH-FIN-Kenya-Project-Brief.pdf>

Also see: Water Sector Trust Fund. (n.d.). Output-Based Aid programme brings affordable piped water to Murang'a South. Available at: <http://www.waterfund.go.ke/stories/oba-muranga-south>

### J.1.3 Recent DFID publications on PbR

Chinfatt, S. & Carson, M. (2017). Supplier access to pre-finance in payment by result contracts. Available at:

[https://assets.publishing.service.gov.uk/media/5a3568fced915d5a62fdc5a5/EXTERNAL\\_REPORT\\_-\\_Supplier\\_Access\\_to\\_Prefinance\\_in\\_Payment\\_by\\_Results\\_Co....pdf](https://assets.publishing.service.gov.uk/media/5a3568fced915d5a62fdc5a5/EXTERNAL_REPORT_-_Supplier_Access_to_Prefinance_in_Payment_by_Results_Co....pdf)

This externally produced report explores the topic pre-financing programmes under payment by results contracts. The evidence presented in this report draws from a literature review on the topic and survey and interview data collected from delivery partners; suppliers; PbR experts and finance providers. By and large, PbR contracts favour larger organisations that have access to cash reserves or qualify for commercial loans; are able to absorb greater financial risk and have robust monitoring and evaluation systems in place necessary under a PbR contract. There are trade-offs for organisations that decide to bid for PbR contracts, as access to finance in any given organisation is limited, particularly in smaller organisations and charities. The results generated in this report also point to other challenges of implementing PbR contracts, for example the delivery context, particularly fragile states and contexts where results are heavily influenced by external actors. Another important challenge that emerged from the data is the issue of setting performance expectations – getting this right is still a challenge.

Clist, P. (2017). Full Report: Review of Payment by Results in DFID: Establishing the Evidence Base: Objective 2: What works for Payment by results Mechanisms in DFID Programmes. Available at:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/684278/full-report-UEA2-merged.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684278/full-report-UEA2-merged.pdf)

The report draws lessons from 20 innovative PbR programmes within DFID's portfolio, including the WASH results Programme. The report provides concise conclusions and recommendations. The reference list also provides a good starting point for updated relevant literature. Interesting aspects that are raised in the recommendations section include differentiating the evidence base between 'big' and 'small' PbR project; applying the emerging evidence base by considering that what may work best in one sector may not equally benefit another sector; and investigating if PbR design and implementation flaws reduce over time.

Clist highlights that the evidence base is still very thin to fully inform how DFID can do payment by results right. In line with DFID's 2014 strategy for payment by results, the results from ongoing evaluations on PbR programmes funded by DFID will become available, doubling the evidence-base for where, how and why PbR mechanisms can lead to greater impact in the short and long-term.

Duvendack, M. (2017). Full Report: Review of Payment by Results in DFID: Establishing the Evidence Base: Semi-systematic review to understand Payment-by-Results mechanisms in developing countries. Available at:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/684277/full-report-UEA1-merged.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684277/full-report-UEA1-merged.pdf)

This review was conducted by an independent team at East Anglia to synthesise the available literature on international development programmes using payment by results. The review identified thousands of studies and narrowed down the pool of relevant literature to 100 research reports based on their selection criteria. The selection used in this review was developed to identify trends and generalizable lessons on how and under what circumstances PbR programmes can achieve their intended outcomes. The review confirms that the majority of available evidence on the topic is within the fields of health and education. The review did confirm a statistically significant positive effect of PbR on targeted results. The reference list in this report will be very useful for the WASH Results synthesis report.

e-Pact Consortium. (2016). DFID WASH Results Programme: Learning Event. e-Pact Consortium. Available at:

[https://assets.publishing.service.gov.uk/media/57a86655ed915d0971000000/Report\\_from\\_DFID\\_WASH\\_Results\\_Programme\\_March\\_2016\\_Learning\\_Event\\_2\\_.pdf](https://assets.publishing.service.gov.uk/media/57a86655ed915d0971000000/Report_from_DFID_WASH_Results_Programme_March_2016_Learning_Event_2_.pdf)

This report was generated following the Learning event that convened DFID, the programme suppliers and the independent monitoring, verification and evaluation team following the end of the Output phase of the WASH Results Programme. This report reiterates much of what has emerged in other reports and literature included in this literature review. Additional topics were raised in discussions, including the importance of including an adequate inception phase with an early payment for Suppliers; a preference to work directly with DFID, as opposed to working with an independent Management Agent and; a strong opposition to working with an independent verification system- systems-based verification is preferred. The report presents the main topics discussed and the main comments from opponents and proponents of each theme. Likely much of this has been captured in the Annexes already developed for the synthesis report.

Feeny, E. (2016). Implementing WASH programmes in a Payment by Results context. Consortium for Sustainable Water, Sanitation and Hygiene in Fragile Contexts. Available at: [https://assets.publishing.service.gov.uk/media/57a3610ae5274a31e0001b61/PbR\\_Learning\\_Brief\\_web.pdf](https://assets.publishing.service.gov.uk/media/57a3610ae5274a31e0001b61/PbR_Learning_Brief_web.pdf)

This short report is published by the SWIFT consortium and provides reflections based on the Consortium's experience as part of DFID's WASH results programme. The report provides insights on how the requirements of a PbR contract can be incorporated into programme design, implementation, monitoring and evaluation.

DFID. (2015). DFID's Evaluation Framework for Payment by Results. Available at: <https://www.gov.uk/government/publications/dfids-evaluation-framework-for-payment-by-results>

This Evaluation Framework outlines DFID's official guidance on how to evaluate Payment by Results programmes. The short report describes the broad priority evaluation questions that should lead to the generation of better evidence to support how PbR mechanisms can improve performance and impact in different circumstances; when and how PbR incentives work in practice. The report touches on different PbR instruments and offers a basis for how better evidence can be generated to assess vfm based on an assessment of evidence on process and impact of the instrument. At a higher level, DFID's Evaluation Framework seeks to provide guidance on how more robust evidence can be generated on the use (both the process and the impacts) of PbR to shape future PbR programming.

Clist, P. & Dercon, S. (2014). 12 principles for payment by results (PbR) in international development. Available at: <https://assets.publishing.service.gov.uk/media/57a089d2e5274a27b20002a5/clist-dercon-PbR.pdf>

In this very brief document, Clist and Dercon produce 12 principles of PbR that have been echoed in DFID's thinking on PbR. The principles listed out attempt to frame PbR not as a silver bullet, but outline the preconditions of its use; the limitations of its use and the possibility of alternative forms of aid that stimulate efficiency gains and improved results.

Clist, P., Verschoor, A. (2014). The Conceptual Basis of Payment by Results. Available at: <https://www.gov.uk/dfid-research-outputs/the-conceptual-basis-of-payment-by-results>

This paper applies two economic models elaborate the conceptual basis for how the costs and benefits compare between aid under PbR and other forms of aid. Analysing the conceptual basis of a PbR programme can indicate if this aid modality is appropriate and cost effective in a given context compared to other forms of aid. The Principle Agent Model is applied alongside the Multitask Model. The Principle Agent Model – focuses on the connection between an outcome that is in the interest of the Principle (e.g., a donor) and payment that is paid to the agent (e.g., an implementing partner). The Multitask Model draws attention to the quality of the performance measure and how this relates to verification and, ultimately determines how strong incentives should be. Clisp and

Verschoor underline that setting performance measures and credible verification systems is a deceptively simple task. The authors work through different measures of appropriateness and present working examples before concluding with a summary of research gaps and main conclusions.

DFID. (2014). Sharpening incentives to perform: DFID's Strategy for Payment by results. Available at:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/323868/Sharpening\\_incentives\\_to\\_perform\\_DFIDs\\_Strategy\\_on\\_Payment\\_by\\_Results.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/323868/Sharpening_incentives_to_perform_DFIDs_Strategy_on_Payment_by_Results.pdf)

This report provides a good summary of the what/why/how of PbR and places this within DFID's longer-term strategy for expanding the use of PbR elements in DFID funding. Why DFID uses PbR? Broadly speaking, the key driver is to improve risk sharing between funder and those who are implementing, which means payments are linked to pre-agreed results, as opposed to milestones or inputs. This mean a shift away from providing funding upfront to fund (certain) activities.

The report also provides different examples of DFID programmes that have a PbR mechanism built into the design. The report also acknowledges that PbR is not suitable for all contexts, especially when results indicators are difficult or too costly to measure. A PbR mechanism can also be integrated in part, only for certain results. It may also be a useful mechanism to integrate into programme design if it is likely to create positive incentives to improve performance; if the benefits of PbR are likely to outweigh the added costs of monitoring, evaluation and verification.

The report suggests a few areas for consideration in deciding how and when to use PbR. PbR mechanisms can be designed by considering 3 main areas: the level of payment on delivery; the type of organisation that is implementing and the type of results that the payment is linked to.

The longer-term DFID strategy for DFID is centred on generating a stronger evidence base for PbR and how it can be tailored to different contexts and expand internal capabilities for doing PbR right. PbR evaluations can explore different thematic areas, including into how the mechanism influences accountability; innovation and flexibility in delivering programme results; increases transparency and accountability; improves performance of implementing partners.

DFID. (2014). Designing and Delivering Payment by Results Programmes: A DFID Smart Guide. Available at:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/352519/Designing-Delivering-PbR-Programmes.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/352519/Designing-Delivering-PbR-Programmes.pdf)

This DFID Smart Guide offers a synthesis of DFID's thinking on PbR, with regards to strategy; insights on designing and implementing PbR, tips and further resources (DFID's PbR Strategy; Clist & Verschoor, 2014 & Clist & Dercon, 2014). The document a good starting point for understanding the promise, preconditions and limitations of PbR. It also summarises DFID's direction of travel away from traditional input financing towards PbR financing, through a mix of different PbR instruments and different implanting agents (Partner governments, suppliers, investors).

Perrin, B (2013). Evaluation of Payment by Results (PbR): Current Approaches, Future Needs. Available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/213938/payment-results-current-approaches-future-needs.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213938/payment-results-current-approaches-future-needs.pdf)



This report was commissioned by DFID and likely fed into DFID's 2014 main PbR strategy document. The report summarises the evidence base of PbR approaches in development and offers a critique of research and evaluation studies that have been conducted on PbR. The report concludes by outlining guidance for future evaluations of PbR. At the time of publication, most PbR evaluations were carried out on PbR implemented in the health sector. The author concludes that the evidence base is weak for how, what contexts and why PbR works. Themes that future evaluations should explore to stimulate the evidence base for using PbR include: counterfactual evidence; incentive mechanisms; PbR cost effectiveness; unintended effects; longer-term impacts and effects of PbR on equity. The authors emphasize that there is a poor understanding of the process of PbR mechanisms and how to manage a PbR project adaptively – to steer when needed and leave flexibility to achieve results.

#### **J.1.4 General search on PbR (2014-2018) in international development assistance in general and in the WASH sector**

USAID. (2018). Scaling Market-Based Sanitation: Desk Review on Market-Based Rural Sanitation Development Programs. Available at:

<https://www.issuelab.org/resources/31021/31021.pdf?download=true>

This report is useful in placing sanitation programmes with a PbR mechanism in the wider landscape of market-based approaches to scaling up sanitation. A variety of case studies are presented in this report following an extensive literature review of sanitation market based interventions. The report discusses case examples involving results based financing (RBF), conditional cash transfers (CCTs) and output based aid (OBA) within the context of other case examples to contextualise how different interventions in the sanitation market have targeted different barriers to demand and supply.

OECD. (2017). Strengthening the results chain: Synthesis of case studies of results-based management by providers. Discussion paper. Available at: <https://www.oecd.org/dac/peer-reviews/results-strengthening-results-chain-discussion-paper.pdf>

This OECD explores seven case studies of development cooperation providers implementing results-based management, specifically focusing on how providers are linking results to accountability, communication, direction and learning. This report analyses how results information is used by development aid agencies along the results chain (inputs, outputs, outcomes, impacts). The OECD suggests that there is an over-emphasis on managing results information for inputs, outputs and outcomes at the expense of longer-term impacts due to greater 'projectisation' of development cooperation, which has contributed to the decoupling of development cooperation results from development results. The OECD outlines six interrelated challenges that a development aid agency must strive to overcome to strengthen the results information that is coupled to longer-term impacts and how results are managed in general. The report concludes with a suite of recommendations that can serve as a point of departure for how development aid agencies can move towards a shared visions for results management.

Global Partnership on Output-Based Aid. (2016). Water Sector Experience of Output-Based Aid. World Bank, Washington, DC. Available at:

<https://openknowledge.worldbank.org/handle/10986/24674>

This report authored by the World Bank's Global Partnership of Output-Based Aid (GPOBA). Based on a review of 12 water supply projects managed by the World Bank (projects only cover piped water supply), the report offers a synthesis of the use and potential of OBA in the water supply sector and lessons learned. The GPOBA's portfolio of water supply projects included in this review suggests that OBA is suitable for water supply



projects in a wide variety of contexts and with a wide variety of implementing agents; OBA has been shown to be able to deliver piped water to poor demographics (household, yard tap and public water point connections). Lessons learned from the water supply projects considered cover a range of lessons from working with utilities and small-scale providers. Lessons included, for example, taking into consideration the capacity of regulators to enforce water supply regulations; striking a balance between the ability of consumers to pay and setting incentives for water suppliers to extend water provision to poorer communities. Unmet challenges included, for example, assessing sustainability of OBA projects; reaching harder to reach communities; how to integrate OBA approaches into larger-scale projects and sector reforms.

UNICEFEAPRO. (2016). Equity in Public Financing of Water, Sanitation and Hygiene (WASH): Analysis from Indonesia, Mongolia and Viet Nam. Available at:  
[https://www.unicef.org/eapro/UNICEF\\_WASH\\_Financing\\_and\\_Synthesis.pdf](https://www.unicef.org/eapro/UNICEF_WASH_Financing_and_Synthesis.pdf)

Results-based financing is discussed in chapter 3 in this report as part of a series of recommendations that UNICEF has proposed to improve subnational financing in WASH. RBF is couched alongside the general recommendation the reforms to public financial management and arrangements for sub-national financing could lead to an enhancement of greater accountability of public officials in improving the quality of sanitation services.

Castalia. (2015). Review of Results-Based Financing Schemes in WASH. Available at:  
[https://www.ircwash.org/sites/default/files/rbf\\_in\\_wash\\_final\\_report\\_full\\_jan\\_2015.pdf](https://www.ircwash.org/sites/default/files/rbf_in_wash_final_report_full_jan_2015.pdf)

The Castalia report stands as the first comprehensive literature review of the use of RBF in the WASH sector specifically. The authors consider 3 types of RBF in the review:

- Output-based aid (OBA): Public or private entities are contracted to provide service delivery where payments, linked to indicators and verified, replace, complement or substitute user contributions
- Conditional Cash Transfers (CCT): Targeted at poor households, where payments are conditional on targeted behavioural changes (e.g. construction of improved household latrines)
- Voucher programmes: Consumers receive redeemable vouchers from government or donor agency, which can be exchanged for goods and services

*Description of projects considered:* Two-thirds of the RBF WASH projects considered in the Castalia report are funded by the World Bank and/or Global Partnership on Output-Based Aid (BPOBA). The remaining funders included unspecified donors, AusAid, USAID, Gates Foundation, Millennium Challenge Corporation and certain State and National governments (Chile, Brazil, Australia), as well as DFID (WASH results programme). Eighty percent of these WASH projects use OBA as the predominant RBF typology, as defined by Castalia in this review. Very few RBF WASH projects considered offered payment for *both* water and sanitation results. At the time of publications, RBF WASH projects implemented in rural areas are more common than in urban areas.

**Key conclusions: RBF projects are effective**, with three quarters of WASH projects included in the review achieving results above target levels. **RBF projects are at least as efficient as conventionally financed WASH projects. There is an absence of data against which to compare RBF WASH projects with conventionally financed WASH projects to determine if RBF is more efficient.** At the time of this publication, no WASH project had published data on the state of service delivery following payment of funds, thus **insufficient evidence is available to indicate that WASH results in PbR projects are more sustainable. This systematic review was limited by the availability of data.** Strengthening the evidence base to assess how and in what ways RBF is a suitable

substitute to more conventional funding for WASH projects requires **comparable data and greater transparency**.

- Available data suggests there may be output bias in reporting. For example, only slightly over half of projects considered reported data on targets and outputs, with most projects only sharing output data.
- A stronger case for project efficiency could be made for RBF projects if comparable benchmarking costs in different contexts for different WASH outputs under RBF and conventional financing were developed.
- WASH projects can integrate RBF to different extents in the design of the overall project; greater transparency is necessary in the separation of data linked to results under RBF and results under conventional payment modes

*Scalability:* At the time of publication, most RBF WASH projects were pilots. Limited evidence suggests that RBF WASH projects can achieve scale if strong local ownership is established and when the RBF modality is well integrated into sector funding arrangements.

*When and where to use RBF?* RBF works well when private providers provide something that beneficiaries' want, when results are measurable and achievable and when the design of the RBF project motivates and incentivises the delivery of results. Limited evidence suggests that public providers perform as well as private providers under a RBF scheme, as the rationale for positive incentives is less obvious with public providers. The use of Conditional Cash Transfers to promote behavioural change has shown limited success so far. Furthermore, there are obvious challenges in verifying behavioural change that occurs in private spaces, for e.g. hand washing and defecation. Greater exploration of innovative monitoring and verification techniques could help improve rigour for monitoring and verification of difficult to measure indicators. More evidence is required to assess the contexts and design considerations where RBF projects can go to scale, be more efficient, effective and sustainable. Of the projects considered, the effect of GDP/capita and government effectiveness of implementation countries had no significant effect on the success of RBF WASH projects, indicating that RBF is a suitable modality in low income and countries with governance challenges.

*Independent third party verification:* Of the RBF projects that presented data on verification and monitoring, the vast majority use some form of independent monitoring. Verification systems need to be designed properly to measure the correct indicators and reduce the risk of gaming and falsification of results leading to payments for results not achieved. Furthermore, if RBF is to lead to more impactful results achieved, then selected indicators, upon which results are verified and payments issued should be linked to outputs (ideally linked to impacts) and not standard project milestones, e.g. 'hygienic latrines built and evidence of use' versus supplier contract signed.

*Impacts and RBF:* Projects that reported impacts, especially health impacts, did not do this convincingly. Linking WASH interventions convincingly to health impacts is a known challenge in the sector and is not solely a challenge for RBF projects (See for e.g., Spears, 2013). Greater transparency of what data is used and how impacts are calculated is necessary to explore the extent to which RBF can lead to greater impacts.

Boyes-Watson, T. (2015). How the increasing use by DFID of Contracts and Payment by results may affect INGOs. Available at: <https://www.mango.org.uk/Pool/N-Mango-Research-into-PbR.pdf>

This report seeks to synthesise how DFID's strategic preference for implementing development cooperation using a PbR approach presents opportunities and particular challenges to INGOs. The author points out that the experience of the UK voluntary sector working under PbR has not been entirely positive, citing polarisation between large- and

small-scale NGOs and overall reduction in financial resilience within the sector due to the shift away from input-based payments. The author presents a series of policy and practical considerations for INGOs when considering signing a PbR contract and a compilation of common problems that INGOs encounter with regards to contract implementation and compliance.

Perakis, R. & Savedoff, W. (2015). Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies. Available at: <https://www.cgdev.org/publication/does-results-based-aid-change-anything-pecuniary-interests-attention-accountability-and>

The authors developed a suite of key features of RBA agreements that can form a basis for predicting programme outcome. The entry points into assessing RBA agreements include looking at: The results level (want a focus on outcomes); Payment function (want payments triggers linked to continuous improvements); Recipient discretion (want greater recipient discretion to stimulate innovation in results delivery); Credibility (want credible funders willing to withhold payments and credible independent verification of results); Transparency (want transparent agreements that make feedback and accountability mechanisms obvious); Payment amount (even small payments linked to results can have a positive effect on outcomes, but smaller payments linked to results may have a lesser impact due to dampening effect on incentive structure). Different RBA programmes are designed to address these features in different manners. The authors suggest that assessing RBA agreements also depends upon the theory upon which they were developed, of which 4 theories are considered (Pecuniary interest, attention, accountability, discretion). The authors assess 4 case studies, which are all RBA agreements with governments, based on the key features outlined above and the underlying theories of RBA effectiveness. The authors conclude that further experimentation of the RBA agreement design will be necessary with governments as aid recipients, as it is more likely that the visibility provided to outcomes under this mechanisms is one of the most distinguishing features from traditional input-based aid.

SIDA. (2015). Results Based Financing Approaches (RBFA) – what are they? Available at: [www.sida.se/contentassets/1b13c3b7a75947a2a4487e2b0f61267c/18235.pdf](http://www.sida.se/contentassets/1b13c3b7a75947a2a4487e2b0f61267c/18235.pdf)

This short report is SIDA's (Swedish International Development Aid) response to their operational staff's demand for guidance on the use of RBFA, particularly how and when it should be used. SIDA is yet to fully develop an official guidance document on RBFA. The report provides an overview of the merits of RBF (A focus on results, transfer of risk to aid recipient, ensuring common goals between donors and aid recipients and increasing aid effectiveness). The report offers a summary of the main variations of RBFAs and summarises the key themes found in the theory of change behind RBFA (financial incentives linked to behaviour change in recipient; funding performance makes results visible and improves management; result-linked payments supports downward accountability to beneficiaries; greater discretion afforded to aid recipients stimulates innovation and adaptive management. The report organises the (pre)conditions alongside questions that SIDA operational staff can ask to assess appropriateness of RBFA. The report concludes with a presentation of examples of how RBFA have been used and brief commentary on results so far. SIDA plans to strengthen the incorporation of RBFAs into SIDA- supported international development cooperation.

O'Donnel M. & Longhurst, R. (2014). Payment by Results: What it means for UK NGOs. Available at: [https://www.bond.org.uk/sites/default/files/resource-documents/bond\\_pbr\\_what\\_it\\_means\\_for\\_ngos\\_nov14.pdf](https://www.bond.org.uk/sites/default/files/resource-documents/bond_pbr_what_it_means_for_ngos_nov14.pdf)

This short report offers a quick and easily-understandable overview of PbR – What is it? What are its aims? What is the evidence base? The report then concludes with a summary a quick summary of advice for NGOs thinking about engaging in a PbR contract and finishes with thoughts on wide issues that PbR could raise in the sector. The report has

been developed with UK NGOs operating under a PbR contract in the UK and within an international development context. The authors warn that the promise of PbR in development is not always achieved, as funders still may restrict flexibility that should be offered to implementing partners to achieve results; PbR agreements should not only have a focus on rewarding results, but should reward learning in complex environments as well. In designing PbR agreements, transparency and downward accountability should be key features. Practically, many NGOs will find that in order to be able to fully engage in a PbR agreement, an upfront investment in capacity building will be necessary to ensure NGOs can manage the increased financial risk and M&E requirements of a PbR contract.

Rodriguez, Diego J., Suardi, M. A., Ham, M., Mimmi, L. M., Goksu, A. J. (2014). Applying results-based financing in water investments (English). Water papers; Water partnership program (WPP). Washington, DC: World Bank Group.

<http://documents.worldbank.org/curated/en/862681468326999086/Applying-results-based-financing-in-water-investments>

The World Bank has followed in the footsteps of other pioneering donors that have increasingly turned to RBF and have experimented with the modality in their contracting. The report cites the UK's DFID, Germany's KfW (Kreditanstalt für Wiederaufbau), Sweden's SIDA (Swedish International Development Agency), and Australia's AusAid (Australian Aid) as the early promoters of RBF. The World Bank has developed a RBF lending programme, Program-for-Results (P4R), and RBF programmes have been developed and supported via the World Bank's International Finance Corporation, as well. The report offers a useful overview of different RBF instruments and offers a clear elaboration of the theoretical underpinnings of RBF; the necessary preconditions for RBF and explores different outcomes of RBF based on actual and theoretical case study examples, although, at the time of publication of this report, much of the RBF experience in the WASH sector had been limited to water supply.

The authors have identified 5 preconditions under which RBF can be successful: (1) Key stakeholders are willing to work with RBF; (2) Implementing agents accept to take on additional risks (particularly financial risk) previously born by the funder; (3) Implementing agents have access to finance from a variety of sources; (4) RBF design and incentive structure is tailored to enabling environment (and may not be appropriate in all environments); (5) Sufficient capacity and competence of funder exists to manage RBF programme.

The authors conclude with an elaboration of 6 principles for tailoring RBF mechanisms to particular contexts. (1) The incentives should focus on the outcome; (2) Indicators should be measurable; (3) Use longer-term perspectives for generating results; (4) RBF programme design should aim to create a market-like or market based system; (5) Ensure allocation of risk is feasible and mitigation measures are in place; (6) Identify & identify potential unintended incentives.

Esseku, H. & Roberts, L. (2013) Development of results-based financing framework for sanitation delivery. Briefing Paper 1770. WEDC International Conference, Nakuru, Kenya. Available at: <https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/30895/2/Esseku-1770.pdf>

The short conference paper discusses the outcomes of a stakeholder meeting that brought together relevant actors that have a role to play in developing access to sanitation in Ghana. This includes relevant government actors from national, regional and district level, development partners and NGOs and community based organisations. A variety of challenges were discussed at the meeting. The authors suggest that insufficient financing and poor incentive structures are the main contributing factor to the poor sanitation outcomes. The authors, thus, propose a RBF framework that outlines outputs and

outcomes, expected achievements and verification process to address the sanitation challenge in Ghana.

UNESCAP. (2013). Development Financing for Tangible results: A Paradigm Shift to Impact Investing and Outcome Models – the Case of Sanitation in Asia. Available at:

<https://www.unescap.org/sites/default/files/Development%20Financing%20for%20Tangible%20Results-A%20Paradigm%20Shift%20to%20Impact%20Investing%20and%20Outcome%20Models.pdf>

This report discusses the notable shift away from input to output towards outcome models for sanitation aid. In this regard, the report provides a good point of comparison between all three development aid models. At the time this report was written, the idea of outcome based aid was in its infancy, but the report suggests that this mode of aid will bring about: greater discipline and opportunities in the marketplace; significantly greater leveraging of private sector and foundation funds; a built-in mechanism integrating social impacts through independent verification and a host of positive externalities achieved through improved sanitation outcomes, which will foster positive incentives and innovative contractual arrangements.

Jenkins, M. & Pedi, D. (n.d.). Guidance Note 8: Equity in Sanitation Marketing: How can we support the market to reach the poorest? UNICEF Sanitation Marketing Learning Series. Available at: [https://www.unicef.org/wash/files/Guidance\\_Note\\_8\\_-\\_Reaching\\_the\\_Poor.pdf](https://www.unicef.org/wash/files/Guidance_Note_8_-_Reaching_the_Poor.pdf)

This UNICEF Guidance Notes presents practical information on how household or supplier out-out based aid can work within the scope of sanitation marketing. Very general, but a quick overview of consumer financing options for sanitation

### **J.1.5 General search on PbR in last 5 years – identifying literature on the use of PbR outside of the development context**

House of Commons Public Administration and Constitutional Affairs Committee. (2018). After Carillion: Public sector outsourcing and contracting – Seventh report of Session 2017-19. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmpubadm/748/748.pdf>

This UK Government report is a direct response to the 2018 collapse and liquidation of Carillion, the UK's 6<sup>th</sup> largest supplier to the public sector in 2017. It is in response to this event that this report was generated. Part of the report includes insights that are relevant to the topic of PbR, even if the context is limited to outsourcing of public services to private and voluntary suppliers under a PbR modality within the UK. The report highlights that the UK Government lacks the ability to assess if a supplier is able to take on greater risk (particularly financial risk). Overloading risk to suppliers risks stifling innovation as suppliers are more likely to play it safe in their approach to delivering results. Governments must offer greater clarity of which risks must remain under the purview of the Government and which risks can be offloaded to suppliers and/or contractors. Clarity on risk management is necessary at the outset, at contracting stage. In this report, The UK Government has been criticised for its poor track record in evaluating results and costing of PbR contracts.

Webster, R. (2016). Payment by results – Lessons from the literature. Available at: <http://russellwebster.com/PbRlitreview.pdf>

This report was commissioned by the OAK Foundation as part of an effort to support decision-making for UK commissioners and providers when deciding whether or not PbR is an effective way to deliver public services. This literature review has limited relevance as it does not explicitly focus on the use of PbR in development. However, the report does provide an overview of the aims of PbR and common criticisms, preconditions for PbR; a



critique of the use of PbR outsourcing within (mostly) the UK context. The report concludes with a suite of conclusions, which caution that the evidence base for the effectiveness of PbR in delivering more effective public services remains weak and commissioners deciding on the appropriateness of a PbR approach have a suite of considerations to take into account, as suggested in this report.

Also see: ICF International. (2015). Payment by results: Learning from the Literature. A review prepared for the National Audit Office. Available at: <https://www.nao.org.uk/wp-content/uploads/2015/06/Payment-by-Results-Learning-from-the-Literature.pdf>

Also see: National Audit Office. (2015). Outcome-based payment schemes: government's use of payment by results. Available at: <https://www.nao.org.uk/wp-content/uploads/2015/06/Outcome-based-payment-schemes-governments-use-of-payment-by-results.pdf>

### J.1.6 List of relevant blogs and newspaper articles identified

Edwards, S. (2018). New results-based financing tool targets clean water for schools. Available at: <https://www.devex.com/news/new-results-based-financing-tool-targets-clean-water-for-schools-92521>

Gustafsson-Wright, E. (2018). From Davos: Is paying for results with blended finance easy to take off? Available at: <https://www.brookings.edu/blog/education-plus-development/2018/02/01/from-davos-is-paying-for-results-with-blended-finance-ready-to-take-off/>

Robinson, Andy. (2018). WASH verification: in vino veritas, in aqua sanitas. Available at: <http://www.snv.org/update/wash-verification-vino-veritas-aqua-sanitas>

Elynn Walter, Guy Howard, Jan Willem Rosenboom, Jeff Albert, Susan Davis, Yi Wei. (2017). Available at: <http://washfunders.org/blog/the-power-of-incentives-lessons-learned-from-designing-and-implementing-results-based-wash-programs/>

Jozwiak, G. (2017). Former DFID political adviser on UK aid direction and his new role at Plan. Available at: <https://www.devex.com/news/former-dfid-political-adviser-on-uk-aid-direction-and-his-new-role-at-plan-90224>

The Economist. (2016). Pay Up: Payment by results is much criticised. But is it wrong to conclude it has failed. Available at: <https://www.economist.com/britain/2016/06/02/pay-up>

Green, D. (2016). Payment by Results take 2: what I learned from the response to last month's rant. Available at: <https://oxfamblogs.org/fp2p/payment-by-results-take-2-what-i-learned-from-the-response-to-last-months-rant/>

Green, D. (2016). Payment by Results hasn't produced much in the way of results, but aid donors are doing it anyway. Why? Available at: <https://oxfamblogs.org/fp2p/payment-by-results-hasnt-produced-much-in-the-way-of-results-but-aid-donors-are-doing-it-anyway-why/>

Hourdin, M. (2016). Vers un financement commercial de l'aide internationale. Available at: <https://blogs.mediapart.fr/manon-hourdin/blog/141016/vers-un-financement-commercial-de-l-aide-internationale>

Lancefield, D. & Gagliardi, C. (2016). What to know before you sign a payment-by-results contract. Available at: <https://hbr.org/2016/09/what-to-know-before-you-sign-a-payment-by-results-contract>

Larsson, N. (2016). Does payment by results dehumanise aid work or encourage success? Oxfam, Swift Consortium and Opportunity International share their experiences of running PbR

programmes. Available at: <https://www.theguardian.com/global-development-professionals-network/2016/apr/25/payment-by-results-dehumanise-aid-work-encourage-success>

Lowe, T. & Evans, K. (2016). Paying for data production: The problem with payment by results. NCVO blogs. Available at: <https://blogs.ncvo.org.uk/2016/06/14/paying-for-data-production-the-problem-with-payment-by-results/>

Webster, R. (2016). The pros and cons of Payment by results. Available at: <http://www.russellwebster.com/the-pros-and-cons-of-payment-by-results/>

Gash, T. (2014). Payment by results fails to encourage smaller suppliers to bid for contracts – Government has tried to encourage bids by smaller companies, but a new report recognises three trends that threaten progress. Available at: <https://www.theguardian.com/public-leaders-network/2014/jan/23/local-gov-lessons-central-gov-suppliers?>

Harvey, A. (2014). OPINION: Securing safe water and good hygiene practices through results-based aid in Uganda. CDKN blog. Available at: [https://cdkn.org/2014/03/promising-preliminary-results-from-results-based-aid-in-securing-safe-water-in-uganda/?loclang=en\\_gb](https://cdkn.org/2014/03/promising-preliminary-results-from-results-based-aid-in-securing-safe-water-in-uganda/?loclang=en_gb)

Bill & Melinda Gates Foundation. (n.d.). Bill Gates Visits Nigeria to Boost Global Fight Against Polio - Bill & Melinda Gates Foundation: Nigerian government to receive more than 100 million doses of oral polio vaccine (OPV) through innovative financing mechanism. Press Room. Available at: <https://www.gatesfoundation.org/Media-Center/Press-Releases/2009/02/Bill-Gates-Visits-Nigeria-to-Boost-Global-Fight-Against-Polio>

Chambers, R. (n.d.). Perverse Payment by Results: frogs in a pot and straitjackets for obstacle courses. IDS Participation, Power and Social Change blog. Available at: <https://participationpower.wordpress.com/2014/09/03/perverse-payment-by-results-frogs-in-a-pot-and-straitjackets-for-obstacle-courses/>

Social Finance. (n.d.) Outcomes Funds. Available at: [https://www.socialfinance.org.uk/sites/default/files/publications/sf\\_outcomes\\_fund\\_note\\_feb\\_2018.pdf](https://www.socialfinance.org.uk/sites/default/files/publications/sf_outcomes_fund_note_feb_2018.pdf)



## Annex K Original Terms of Reference

### K.1 Original TOR

#### 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<sup>53</sup> 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

<sup>53</sup> 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.

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](http://www.gov.uk/DFID)

10. DFID' approach<sup>54</sup> 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 contracts
- Un-announced visits to project work.
- Quality assurance of contract holders results returns

<sup>54</sup> <https://www.gov.uk/government/publications/payment-by-results>

- 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<sup>55</sup> 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;

<sup>55</sup> <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

- a quality assurance plan;
- an outline of proposed methods for assessing core indicators;
- 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 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 (e.g. 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 improved sanitation facilities). DFID has provided further information<sup>56</sup> 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.

<sup>56</sup> <https://www.gov.uk/government/publications/dfids-approach-to-value-for-money-vfm>

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.
- 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.

## K.2 Clarifications to the TOR

This section clarifies our understanding of the TOR for this evaluation. The TOR 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 J.1 for the full TOR).

We propose to address a selection of these issues through activities across two interlinked workstreams – namely the **Programme evaluation Workstream (PEW)** and the **RCT Research Study**. Based on the list outlined in the TOR, 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 TOR 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 learning and dissemination 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 TOR outlined that the evaluation team would select, design, and administer one or more in-depth evaluations at programme level and 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 OECD DAC criteria situate '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 been reduced by a certain percentage, but we cannot explain

why it was not reduced by more than this; or we may 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.
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 the 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 TOR:

- *‘identifying factors which have enhanced or impeded the sustainability of WASH interventions’;* and
- *‘assessing how the wider operating environment has enabled or impeded 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 TOR:

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

We understand the former to be part of the tracking of value for money (VFM) metrics, listed under *verification* activities in the TOR. However, the TOR 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 TOR 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 high-level evaluation questions (HEQs) along the DAC criteria, which were signed off by DFID in September 2014.