Strengthening evaluation effectiveness and impact

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

Evaluation Synthesis Report – Executive summary

January 2019

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.

Acknowledgements

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

We would like to thank Guy Howard, Stephen Lindley-Jones, Leonard Tedd, and Laura Westcott at the UK Department for International Development, the suppliers and their partners (especially their teams in Kenya, Pakistan, Tanzania, and Uganda), and the monitoring and verification team for their time and inputs during this period.

The core evaluation team comprised Julia Larkin, Lucrezia Tincani, Zach White, and Jeremy Colin. This core team worked with a wider team in conducing the data collection and analysis; this wider team included: Sue Cavill, Sarah Javeed, Alice Mango, Faith Muniale, and Shona Jenkins. The following experts served as internal expert reviewers: Alex Hurrell, Sophie Witter, Richard Carter and Timothy Forster.

The draft report was also reviewed by a representative from each supplier (John Dean from SAWRP, Anne Mutta from SSH4A and Joanna Trevor from SWIFT) as well as representatives from the Itad verification team (Katherine Welle and Ben Harris).

The evaluation findings draw on hundreds of interviews conducted with consortium programme staff, local implementing NGOs, and government counterparts. The evaluation team are extremely grateful to all those who offered their time and insights.

Disclaimer

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). 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). The client reference number for the project is PO6507.

e-Pact
Registered in England: 3122495

Level 3 Clarendon Centre
52 Cornmarket St
Oxford OX1 3HJ
United Kingdom

Tel +44 (0) 1865 207300
Fax +44 (0) 1865 207301
Email admin@opml.co.uk
Website www.opml.co.uk

Executive summary

The WASH Results Programme

The UK Department for International Development's (DFID) Water, Sanitation and Hygiene (WASH) Results Programme is a £111 million seven-year programme (2014 to 2021). The first round of the programme was implemented between May 2014 and March 2018 and aimed to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people in 11 countries[1]. In 2017 the programme was extended, to include a second phase lasting until 2021. The scope of this evaluation is limited to the first contracting round (2014 to 2018)[2].

The programme was implemented by three supplier consortia contracted to DFID under a payment by results (PbR) financing modality. The WASH Results Programme was one of DFID's - and the wider sector's - first large-scale applications of a PBR contracting modality. Suppliers had no up-front financing; 100% of payments to Suppliers were based on delivery of pre-specified 'results' which were independently verified by a third party.

The nature of the 'results' varied by supplier and varied across the phases of the programme. In the two initial years of the programme (the 'output phase' – January 2014 to December 2015) payments were made for the completion of activities and output-level results. While in the latter two years of the programme (the 'outcome phase' – January 2016 to March 2018), payments were linked almost exclusively to achievement of outcomes, with a small proportion of payments linked to indicators of sustainability of outcomes for one Supplier.

In addition to the PbR financing modality the WASH Results Programme had a number of other key features, the most important of which were:

- **size**, 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 delivered access to water and sanitation services and hygiene promotion campaigns; and an outcome phase (to March 2018), which aimed to ensure the sustainability of the services and behaviour changes achieved.

The Suppliers

The programme was implemented predominantly through INGOs working in partnership with local government and national NGOs. Two suppliers, the Consortium for Sustainable WASH in Fragile Contexts (SWIFT), and the South Asia WASH Results Programme (SAWRP), were consortia of international non-governmental organisations (INGOs). The third, Sustainable Sanitation and Hygiene for All (SSH4A), was implemented by a single INGO. Table 1 (overleaf) provides an overview of the three consortia.

[1] South Sudan and Liberia were dropped from the original 13-country programme
[2] Envisaged as a £70 million four-year programme (May 2014 to March 2018)

Table 1: Consortium composition

Programme and budget	Lead supplier and consortium partners	Countries of operation
SAWRP ***£24,995,906***	**Lead:** Plan International **Consortium partners:** WaterAid, Water Sanitation for the Urban Poor (WSUP) and Unilever. Water Engineering and Development Centre (WEDC) and Ipsos MORI **Implementing partners**: various national NGOs	Pakistan, Bangladesh
SSH4A ***£24,032,477***	**Lead**: SNV **Consortium partners:** none **Implementing partners:** various national NGOs and government partners	Ghana, Ethiopia, Tanzania, Uganda, Kenya, Mozambique, Zambia, Nepal ***Dropped****: South Sudan in 2016 due to fragile-state context*
SWIFT ***£19,668,078***	**Lead:** Oxfam **Consortium Partners**: Tearfund, WSUP; the Overseas Development Institute (ODI) **Implementing partners:** Practical Action, Sanergy, Concern Worldwide, and various national NGOs and public water utilities	Democratic Republic of the Congo (DRC), Kenya. ***Dropped:*** *Liberia in 2014 due to the Ebola crisis)*

The SWIFT consortium was led by Oxfam GB and operated in the Democratic Republic of the Congo (**DRC) and Kenya**. In contrast to the other supplier programmes, activities encompassed both urban and rural contexts. SWIFT was the only consortium that included a substantial water supply component in addition to sanitation and hygiene promotion. In DRC, SWIFT delivered water and sanitation and hygiene activities at scale though the national government WASH programme. SWIFT in Kenya operated through a portfolio of small- to medium-sized projects across both urban and rural sanitation and water; with implementation through INGOs.

The SAWRP consortium was led by Plan International and operated in Bangladesh and Pakistan. The focus of SAWRP was on rural sanitation and hygiene, though it included a modest[3] water supply component. SAWRP implementation was through community-level programming and implementation in schools. Plan International, WaterAid, and their local implementing partners promoted sanitation and handwashing with soap in poor rural communities using community-led total sanitation (CLTS) and sanitation marketing. Unilever delivered a handwashing campaign in a large number of schools in both countries, though the focus was overwhelmingly in Bangladesh.

SSH4A was solely implemented by SNV and its local implementing partners in seven African countries and Nepal. The SSH4A programme focused only on rural sanitation and hygiene promotion– it had no water supply component. Activities were implemented within the framework of the existing SSH4A programme, which has been operating in a dozen countries in Asia since 2009. The SSH4A approach embodies a tested operational model for working at scale and combines demand creation − mostly using CLTS − with support to sanitation supply chains, behaviour change communication (BCC), and strengthening capacity for WASH governance. Implementation at the field level was predominantly by local government partners, supported to a greater or lesser extent by NGOs (referred to as Local Capacity Builders (LCBs) by SNV).

[3] Water supply improvements were not a major feature of the programme (only 3% of beneficiaries), but were included in order to provide access to some of the poorest and most under-served poor communities.

The verification approach

The independent verification of suppliers' results adopted a systems-based approach, meaning it was based on data generated by the supplier's internal monitoring and reporting systems, rather than independent data collection by the monitoring and verification (MV) team. Working with DFID and suppliers, the MV team created a verification system for the programme – tailored specifically to each supplier's monitoring and reporting systems. However the verification approach contained the same core elements:

- **Systems appraisals** of the suppliers' monitoring and evaluation (M&E) systems were conducted to assess the likelihood that data would be reliable and reports would be timely and accurate.
- **Field visits** were conducted to better understand supplier M&E systems, to collect primary data that feeds into verification, and to help build relationships with Suppliers.
- **The evidence submitted by suppliers was reviewed,** using a set of minimum standards. Sometimes spot checks were done on the data.
- **After Action reviews were held** to ensure that key learning surrounding measurement and verification of indicators was applied in subsequent verification rounds.

Programme performance

By most measures the programme was very successful in achieving its stated objectives. The vast majority of output-level targets were achieved by all suppliers, and in many areas with significant overachievement. A prominent reason for exceeding targets was that suppliers had planned for overachievement as part of their risk management strategy. The suppliers also overwhelmingly achieved the outcome targets, with significant overachievement in several areas and modest underachievement in relatively few others. The WASH Results Programme also consistently scored well under DFID's Annual review process; scoring A in 2014 and 2015 and A+ 2016-2018.

The evaluation approach

The WASH Results Programme evaluation was primarily undertaken for learning purposes, not accountability.[4] The objectives of the evaluation were 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.

The main focus of the evaluation, and this report, is on how and why and under what circumstances results were achieved. The 27 detailed evaluation questions address all five Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria: relevance, effectiveness, efficiency, impacts, and sustainability.

[4] Operationalisation of the original terms of reference resulted in two distinct workstreams. The Programme Evaluation Workstream (PEW) is addressed in this document. The randomised control trial (RCT), conducted in Pakistan, which examines factors affecting the sustainability of sanitation behaviour change in Pakistan, is described in separate documents.

The mid-term evaluation report (2016) examined progress half way through the programme, focusing particularly on the original design and tendering process, and on the output phase (March 2014 to 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.

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.

This final synthesis report incorporates key findings from the midline with the results from the new endline evaluation activities. The findings draw on an analysis of programme documentation; verification data; an analysis of the 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.

Lessons learned

The lessons presented in this section are framed around the core learning objectives of this evaluation as opposed to specific evaluation questions. As an endline evaluation, the evaluators focused on broader lessons relevant for the future for individual suppliers, DFID and/or the WASH sector as a whole. This report does not repeat all of the lessons and recommendations from the midline evaluation (see full list from the midline in Annex A), though several themes carry through. The lessons presented here are further elaborated upon and nuanced in the main body of the report.

(i) Whether the programme successfully achieved its stated objectives

A. **All three supplier consortia were successful in delivering their targets, through there is a question of how replicable this achievement is**. Suppliers achieved, if not exceeded, the vast majority of their contractual targets; partially due to the possibility to shift over-achievement to cover underperformance within some subcategories. However, the tight timetable in the output phase placed a high burden on staff; particularly during the output phase. The WASH Results Programme was exceptional in that it was the first large-scale application of PbR in the WASH sector; it was a high-profile programme which presented great financial and reputational risks for the suppliers. With this came heightened management attention. These exceptional factors are significant because they raise some questions surrounding how replicable the achievements of the programme may be.

B. **Suppliers' programme approaches were generally well aligned with DFID's expectations, but in some cases the fact that DFID did not articulate these expectations more clearly resulted in a missed opportunity.** The programmatic approaches chosen by the suppliers generally met DFID's expectations, such as operating at scale[5], inclusivity, and sustainability. However, not requiring a focus on community-wide coverage (in sanitation Open Defecation Free (ODF) status) or other best practice for achieving sustainable and inclusive outcomes presented a missed opportunity – though it is noted that in many cases ODF was still pursued as a programme objective by suppliers. DFID not explicitly setting all of their expectations at the bidding stage created a risk for potential suppliers, onto whom the risk of delivery was transferred under a PbR contract.

[5] Taken here to mean reaching a large number of people.

C. **The PbR modality strengthened supplier monitoring systems, though in many cases undertaking improvements to monitoring systems was more burdensome than suppliers anticipated.** The monitoring systems of most of the supplier consortia partners required moderate to significant modifications to meet verification requirements. It is important to strike a balance between increased rigour and the corresponding increased cost, because there is a potential opportunity cost with regards to staff time and money in doing ever more intensive monitoring.

D. **Suppliers reported substantial learning on how to manage and price risk**. Based on the programme's experience, the following approaches worked well, some of which can be considered normal good practice: (1) setting realistic targets; (2) using reliable implementation partners; (3) applying tested approaches; (4) including a contingency fund; (5) considering cost inflation in the total price; (6) considering the risk that partners would not achieve the intended results; and (7) taking advantage of the increased monitoring to facilitate risk management.

(ii) The influence of programme design, including the PbR modality, on this achievement

E. **DFID gave suppliers substantial flexibility and freedom to design and implement their own approaches, but this flexibility largely manifested only to higher levels of programme management**[6]. Suppliers were able to propose the indicators, and the frequency with which they would be assessed, and each supplier individually negotiated a different 'results framework' with DFID under which they were paid. Suppliers appreciated this flexibility at the design stage, as well as the flexibility allowed during implementation. This flexibility was largely confined to higher levels of programme management, as field teams were tightly managed, and where there were local implementing partners they generally operated on grant agreements which retained financial and activity reporting.

F. **Implementation approaches for the three supplier consortia varied widely yet overwhelmingly relied on proven approaches and existing relationships with NGOs and governments**. The evaluators uncovered little evidence of genuine innovation in programmatic approaches – meaning that no approaches that were novel to the global WASH sector were developed, though it is important to note evaluators found several instances of adaptive programming.

G. **The split between output and outcome phases had mixed results.** The urgency of the December 2015 deadline for the output phase in some cases limited suppliers' ability to adapt and posed risks to quality, at least in the short term, and the sustainability of results. While it was clearly beneficial to have longer-term engagement with the institutions (government or community-level) with responsibility for managing services, this could have been achieved without formal phasing. Similarly, while imposing accountability for outcomes was beneficial and improved functionality, this could also have been achieved without formal phasing.

[6] This lesson is also supported by evidence from the wider PbR literature. See Soucat, A., Dale, E., Mathauer, I., Kutzin, J. (2017) 'Pay-for-Performance Debate: Not Seeing the Forest for the Trees'. *Health Systems & Reform* 2017; 3:74–79; and Renmans, D., Holvoet, N., Orach, C. G. and Criel, B. (2016) 'Opening the "black box" of performance-based financing in low- and lower middle-income countries: a review of the literature'. *Health Policy and Planning*, 31(9), 1297–1309. doi: 10.1093/heapol/czw045

(iii) Lessons for applying PbR in WASH programming in future

H. **The WASH Results Programme was not a clean test of PbR**, for a variety of reasons, such as the evolution of expectations, the characteristics of the contracting process, and the uniquely tight timetable for the first two years. For example, many programmatic decisions had already been made by suppliers, by verifiers and by DFID before the PbR modality was fully understood, and before the indicators for which suppliers would be judged were negotiated. While the WASH Results Programme offers several interesting and indicative lessons, its contribution to the broader evidence base for PbR in WASH is limited.

I. **PbR appeared to be a viable option for funding WASH programmes, but 100% PbR was neither necessary nor optimal.** To help manage cash flow and other risks, two suppliers selected several process or activity targets, however these activities were time-consuming to document and verify. These could easily have been reimbursed in a more traditional way, without reducing the overall stimulus a PbR contract provides.

J. **A PbR approach using disincentives only, without bonus opportunities, put potentially undue, and unintended, burden on suppliers.** Suppliers put in a great deal of effort to not only achieve but overachieve many targets, but were not rewarded for doing so.

K. **The efforts of suppliers were not solely informed by financial incentives.** The suppliers contracted under the WASH Results Programme had strong organisational values, which led them to take action on aspects not directly linked to payment. In many cases the reputational risk associated with not meeting given targets was also a powerful motivator.

L. **Clarity on guiding principles of how major events, such as natural disasters or epidemics, would be handled between DFID and suppliers is important in a PbR context.** In two cases involving SWIFT, DFID *de facto* assumed some of the financial risk after events. As the level of risk-sharing and the level of evidence expected was not clearly specified at contract stage, time and resources were invested in negotiating the level of risk sharing.

M. **The high level of uncertainty at invitation to tender stage discouraged several bidders.** Specifically, the lack of clarity surrounding the results and how they would be verified meant that some private sector actors viewed the programme as very high risk.

N. **Under the WASH Results Programme, in part because of the lack of clarity on how the PbR modality would operate, the PbR incentives acted within the context of tried and tested implementation approaches than on untested operational models.**

O. **The lack of an inception phase – to design verification requirements before implementation began – caused inefficiencies because the verification had to be repeatedly adapted as the supplier's systems changed or became better-known by the MV team.**

P. **The value of including process-related indicators as payment indicators was dependent on context.** The evaluators perceived that the process indicators related to start-up activities were included primarily to minimise the risk of non-payment for the suppliers and to facilitate cash flow in early stages rather than incentivise supplier attention on specific aspects of programming[7]. This was suboptimal and inefficient when these indicators are costly and time-consuming to document and verify. The experience of this

[7] This indicates that it is less likely that such process indicators would be included in future if a significant part of the programme was grant-funded.

programme shows that more complex aspects of programming (e.g. sustainability prerequisites or learning) which lack agreed standards are difficult to measure. There is value in reporting on process indicators related to these complex aspects of programming when DFID seeks to encourage certain aspects in programme implementation. Though this reporting does not necessarily need to be linked to result-based payment[8].

Q. The PbR mechanism was better able to incentivise the desired results in those cases where suppliers had more control over results. The PbR incentive acted less strongly where implementation was through government, and where activities were incentivising behaviour change.

R. Larger international partners were better able to handle the pre-financing, risk management, and flexibility necessary under PbR than small organisations. All three lead suppliers chose to hold a large portion of risk centrally. There were mixed results in the few cases where local implementers were on full or partial PbR contracts, with some shifting positively to a stronger sense of accountability, while others were less able to cope and in some cases had their contracts discontinued.

S. Establishing appropriate outcome level targets was challenging. This was because much was beyond the suppliers' direct control, and at the time of contracting there were few established benchmarks for the conversion of WASH outputs into outcomes.

T. Several lessons emerged on how the PbR modality could best strengthen supplier monitoring systems: In the WASH Results Programme, explicitly linking payments to outcomes triggered intensive discussion around outcome measurement, and the validity of those measurements in particular. Outcome-level monitoring, as well as accountability, played a facilitative role in ensuring functionality. The increased monitoring also supported supplier risk management strategies. It is also noted that suppliers were generally positive about the MV's contribution to strengthening monitoring.

U. Several missed opportunities for learning occurred under the WASH Results Programme: data were primarily used for progress reporting/compliance, as opposed to learning; and the tight schedule in the output phase also limited opportunity for deeper learning and reflection.

[8] If aspects such as inclusion were linked to a result-based payment, these payments could for example reward the following: (a) Work in under-served geographical areas could be rewarded by allowing a higher unit price-per-beneficiary. (b) Bonus payments could reward survey results which confirm that equity targets and water-point functionality targets have been met after a certain period.

Recommendations for future WASH PbR programmes

These recommendations are primarily for donors, such as DFID, that are considering the use of PbR in future WASH programmes.

1. **At design stage clarify the purpose of using PbR and consider the implications of PbR for the type of supplier expected to bid.** While the *potential* benefits of PbR generally are well documented, it is important to set out in the programme design what the *specific* rationale is for using PbR in this case, and to define the added value that it should bring.

2. **100% PbR may be undesirable – the size of the PbR component should be tailored to the purpose specified and to the feasibility of measuring results** – especially in cases where suppliers require finance in the early stages of implementation to ensure cash flow. We recommend a hybrid design where a part of supplier payments are grant-based, with a smaller percentage used as an incentive for good performance in key areas[9]. If PbR is used to incentivise action related to aspects of programming that are important but difficult to measure[10], then great care should be given to the indicators used

3. **Where possible, streamline the verification burden on suppliers.** This relates both to the indicators used for payment purposes and how the effort is applied by verifiers.

4. **At design stage, the funding agency should, as far as possible, provide more clarity on the results to be achieved and the accompanying verification requirements.** This may include appointing the verification provider before implementation begins and/ or being more prescriptive on standards for verification requirements[11].

5. **As far as possible, at tender stage clarify donor/supplier risk-sharing arrangements in the event of exceptional events** including the level and type of evidence expected.

6. **In future programmes ensure there is a sufficient inception phase,** ensuring that verification requirements are clear before any implementation activities begins.

7. **Ensure that the learning by the MV team and suppliers on measuring and verifying key WASH indicators (especially outcome-level indicators) is captured and disseminated at the sector level.**

[9] Experience with PbR in other sectors also suggests that the value of PbR lies not in reimbursing the direct costs of implementation (as NGOs are already motivated to do this) but in linking PbR payments so as to incentivise only those dimensions which might otherwise be marginalised, albeit unintentionally

[10] These aspects include: creating an enabling environment for sustainability; ensuring meaningful action on equity and inclusion; and allowing space for learning.

[11] This recommendation is also supported by evidence from the wider PbR literature. See Witter, S. *et al*. (2018) '(How) does RBF strengthen strategic purchasing of health care? Comparing the experience of Uganda, Zimbabwe and the Democratic Republic of the Congo'. Submitted for publication.