

Research Paper

Does payment by results work? Lessons from a multi-country WASH programme

Guy Howard and Zach White

ABSTRACT

Payment by results (PbR) for financing public services has attracted increasing interest over recent years in the water, sanitation, and hygiene (WASH) sector. PbR is attractive to funders as a mechanism because it focuses attention on results rather than inputs, and because it transfers a proportion of risk to suppliers. This paper reviews the experience of the UK Department for International Development (DFID) funded WASH Results Programme (WRP), which used PbR, drawing on a process evaluation and the experience of the first author in commissioning the programme, and the second author in evaluating it. The WRP met its targets for people reached with first-time access to water and sanitation and generated high-quality programme data. The PbR mechanism provided strong incentives to the suppliers to improve their monitoring systems. However, the suppliers tended to use tried and tested approaches, with limited innovation. It is critical to consider certain key elements in the design of PbR programmes, including the proportion of funding that uses PbR and the proportions of PbR that focuses on outputs and outcomes.

Key words | payment-by-results, finance, results-based finance, WASH results, WASH targets

INTRODUCTION

Payment by results (PbR) is an approach to financing public services that has gained increasing attention over recent years (Cabinet Office 2011). PbR is a form of outcome-based performance management that provides financial rewards for the achievement of specific outcomes (Albertson *et al.* 2018a). PbR and other approaches to outcome-based performance management are derived from the sustained efforts of governments, particularly in high-income countries, to improve efficiency and effectiveness in service provision as part of reforming public services, which has been referred to as the 'New Public Management' model (Hood 1991).

The use of PbR is attractive to funders of public services because it places a focus on achieving a desired set of results,

rather than managing a range of inputs. It transfers a portion of the delivery risk to suppliers, which is accompanied by increased responsibility of suppliers for delivery (National Audit Office 2015). In a review of the drivers for developing PbR, Albertson *et al.* (2018b) note that there may be several objectives for PbR, including expansion of the supplier market and diversifying the supplier base; improving efficiency in public service delivery; and managing complex issues with multiple factors that influence outcomes. Clist (2017) notes that the key aspect for PbR is the quality of the performance measure and that in order to create the right incentives such measures must effectively capture a result that all those involved in a project really care about. In the context of water, sanitation, and hygiene (WASH), this includes expanded access to safe drinking water and/or sustained use of services.

From the beginning of 2010, the UK Government placed an increasing focus on expanding the use of PbR in ODA-spending

Guy Howard (corresponding author)
Department of Civil Engineering,
University of Bristol, University Walk,
Bristol BS8 1TR,
UK
E-mail: guy.howard@bristol.ac.uk

Zach White
Water, Sanitation, and Hygiene Team,
Oxford Policy Management,
Clarendon House, Level 3, 52 Cornmarket Street,
Oxford, OX1 3HJ,
UK

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/washdev.2020.039

through the Department for International Development (DFID). By the end of 2013, 71% of centrally issued DFID contracts had a performance-based element. In 2014, DFID published its strategy on PbR, which included commitments to make PbR its 'business as usual' approach in contracts with its suppliers, rather than the exception (DFID 2014b).

The evidence base from evaluations of PbR programmes in the context of official development assistance (ODA) remains limited. Clist (2017) and Fox & Morris (2019) note that there remain few evaluations of PbR programmes and that the quality of evaluations is generally poor. Albertson *et al.* (2018b) summarised that the findings of the evaluations that were available, primarily based on examples from the UK. They found that PbR has often associated with more complex commissioning and increased commissioning costs; mixed results in expanding supplier markets; an increased focus on performance, often through strengthened monitoring, which had both positive and negative effects; increases in delivery costs in some cases; no evidence of extensive innovation and that suppliers were risk-averse; and limited evidence of 'gaming' the PbR system.

PbR programmes can deploy downside and upside incentives to the suppliers contracted. Downside incentives mean that failure to meet expected targets results in a loss of payment. Upside incentives reward suppliers with additional payments should the desired results be exceeded. Clist (2017) concluded that a key element for the success of PbR is the ability and willingness of the funder to withhold payment, especially in relation to non-governmental organisations (NGOs). This focus on downside incentives, however, tends to make suppliers risk-averse and to stifle innovation.

This paper reviews the experience of one WASH programme that used PbR – the WASH Results Programme (WRP) – to examine how PbR has worked in practice, and the lessons learnt. The paper examines the key question of whether the programme delivered the expected results for the funder of the programme. The paper does not assess whether PbR worked *better* than other approaches to funding WASH. There are insufficient data to make a fair comparison as data were not collected from comparable programmes that had the same set of target results, similar geographies and similar timeframes, which would be required to make a comparative analysis.

THE WASH RESULTS PROGRAMME

The WRP is a £111 million programme running from 2013 to 2022. It was originally designed to meet a UK Government target for numbers of people gaining access to WASH between 2010 and 2015, but further extended in 2016 to support the achievement of greater numbers of people gaining first-time access to WASH between 2015 and 2020.

The WRP is delivered through three supplier contracts: the South Asia WASH results programme (SAWRP) consortium, led by Plan International; the Sustainable WASH in Fragile Contexts (SWIFT) consortium, led by Oxfam; and the SNV Sustainable Sanitation and Hygiene for All (SSH4A) programme. The contracts include two distinct sets of results associated with two phases of activity: an output phase, focused on ensuring first-time access to sanitation and/or water supply supported with hygiene education and a subsequent outcome phase to maintain use of sanitation and/or water supplies constructed and to maintain good hygiene practice.

Table 1 summarises the suppliers' programmes and the number of people expected to be reached. The WRP achieved, and in most cases exceeded, the results targets established. At the end of the output phase under the original contracts, the WRP projects had provided over 1 million people with first-time access to water supply, over 4 million people with first-time access to sanitation, and over 10 million people reached with hygiene messages (DFID, 2016). The outcome phase focused on continued use of services as shown in Table 1.

At the start of the programme, an independent team of sector experts were contracted to provide verification that the results claimed by suppliers had been achieved, and therefore, payments could be made. Verification was primarily based on systems appraisals, assessing whether supplier monitoring and data systems were robust and reliable and, therefore, producing credible data, with limited spot-checks in the field. An autonomous team undertook both process and impact evaluations.

The WRP only used downside incentives, operating on a sliding scale down to 70% of the target, after which full payment would be lost. The WRP was designed as

Table 1 | Details of supplier targets and countries included in programmes and outcome results from the original contracts (DFID 2018)

Supplier	Total target people to be reached	Countries	Outcome results
SAWRP	2,279,761	Bangladesh, Pakistan ^a	Bangladesh: 96.1% of people continued to use latrines and 95% continued to use water supplies Pakistan: 99% of people continued to use latrines and 100% continued to use water supplies
SSH4A	3,471,797	Bangladesh, Nepal, Ethiopia, Ghana ^a , Kenya, Mozambique, South Sudan ^b , Uganda, Zambia	2.85 million people continued to use latrines (exceeded target by 30%)
SWIFT	1,684,402	Democratic Republic of Congo (DRC), Kenya ^a , Liberia ^b	70% of people continued to use latrines and 75% of people continued to use water supplies

^aCountry not included in contract extension.

^bCountry where activities were stopped early.

'100% PbR' – that is, all payments were contingent on the achievement of a pre-defined set of results. However, payments were not solely linked to output- and outcome-level 'results': interim results were set within each phase to allow payments to be made throughout the programme. Many early payment triggers were essential programme activities and inputs (e.g. training workshops).

METHODS

This paper draws on data collected through a process evaluation of the WRP (ePact 2018) and complements this with an analysis of the experience of the first author in the commissioning and management of the programme. We draw these two forms of evidence together through a critical analysis with greater weight given to empirical data over expert experience. The analysis of the findings of the evaluation and expert opinion are structured to primarily address the needs of the funder, as the data collection was designed to meet this purpose.

The data collected through the evaluation included: key informant interviews with DFID management, the global management of the three supplier consortia; interviews with programme management in the 11 WRP countries; and case studies in 4 of the 11 countries, which included interviews with field-level staff, government counterparts, and beneficiaries. The primary data were collected over the course of the evaluation, supplemented by a literature

review, a review of programme documentation, and reviews of the programme annual reviews and business cases.

This evaluation drew on elements of contribution analysis and realist evaluation to assess the degree to which the PbR modality influenced implementation. This paper focuses on one of the core hypotheses explored in the evaluation. Namely, that the introduction of a PbR modality helped to achieve intended outputs and outcomes. This assertion is based on three related propositions:

1. the programme and its PbR modality allowed the flexibility of implementation approach within the sub-programmes, which helped to achieve output and outcome objectives;
2. stronger monitoring systems as a result of the PbR modality increased the likelihood of achieving intended outputs and outcomes; and
3. the results-oriented problem-solving promoted under the PbR modality increased the likelihood of achieving intended outputs and outcomes.

The analysis presented here does not cover the full breadth of the evaluation but provides a more detailed assessment on these three critical propositions and explores key issues of interest to funders seeking to use PbR. This includes the degree to which the PbR modality helped to achieve DFID's stated market-shaping objectives.

There are limitations to the study presented here, most importantly that the data collected and analysed does not permit a comparative assessment of PbR performance for

the reasons noted above, and that detailed value for money analysis was not possible. In addition, the focus on the needs of the funder, means that the experience and impact on suppliers and communities, is not fully captured.

RESULTS

Did the PbR modality allow the flexibility of implementation approach within the sub-programmes, which in turn helped to achieve output and outcome objectives?

While there was flexibility in how suppliers met their targets, there is little evidence of innovation. There were several important reasons for this, including that many of the design decisions were made by partners before the PbR modality was fully understood. Furthermore, in the face of high downside delivery risks, the partners generally adopted tried and tested approaches in contexts familiar to them, as these could predictably deliver results. The main flexibility that supported the achievement of the targets was the ability of the suppliers to use multiple projects to deliver results, meaning that shortfalls in one country could be offset by achievements in other countries.

The evaluation found that the removal of financial and activity reporting requirements did result in programme managers being able to more flexibly manage programmes, but that this flexibility tended to stay at the higher levels of programme management. There were two principal reasons for this. First, many consortium leads were unwilling to transfer that level of risk to partners as they did not believe they had the ability to pre-finance activities. Secondly, there was a perceived need to tightly manage partners and field teams to ensure results were delivered, and complete financial autonomy at that lower level was seen as too risky.

Did stronger monitoring systems as a result of the PbR modality increase the likelihood of achieving intended outputs and outcomes?

The systems-based approach to verification under the WRP provided significant incentives to the suppliers to improve their monitoring systems and the suppliers made substantial

investments to improve the reliability and robustness of their monitoring systems, which led to more robust data. These improvements were sustained in the 'outcome phase' following the delivery of outputs. The evaluation found that the suppliers underestimated the amount of time, effort, and resources required to upgrade their monitoring. Given that PbR was unfamiliar to suppliers and the DFID team, the evaluation concluded that an inception phase would have been beneficial to allow for the design of verification before implementation. The WRP suppliers and verifiers iterated successfully around verification processes, but this was inefficient and could have been avoided by having greater clarity earlier in the programme.

Verification of the more complex aspects of programming (e.g. sustainability prerequisites or learning) lacked agreed sector-wide standards for measurement proved difficult to measure. Furthermore, establishing appropriate outcome-level targets was challenging because attribution was more difficult, and because of the limited evidence of suitable benchmarks for the conversion of WASH outputs into outcomes.

Did the results-oriented problem-solving promoted under the PbR modality increase the likelihood of achieving intended outputs and outcomes?

The achievement of results targets within the expected timeframe, which was challenging both in terms of numbers of people and limited implementation time, demonstrated that the PbR modality using strong performance measures, created positive incentives for a focus on results delivery. However, the evaluation of the WRP found that the approach of linking all payments to results was 'neither necessary nor optimal'. The application of PbR in the early stages of the programme meant the results being verified were process-related indicators primarily used to facilitate cash flow, rather than to incentivise supplier attention to specific aspects of programming. This was inefficient as these indicators were costly and time-consuming to document and verify.

Over the course of the programme, three external shocks severely impacted on programme results. In each case, these events led to negotiations surrounding the appropriate level of supplier–DFID risk-sharing as the results

could not be fully achieved as originally intended. These events were the Ebola outbreak in West Africa in 2014, a resurgence of violence in South Sudan in 2016, and a prolonged drought in East Africa in 2017–2018. In each of these cases, DFID *de facto* assumed some of the delivery risk *post hoc*. In the case of Ebola, the contract was renegotiated to exclude affected countries, with adjustments upwards in price per beneficiary made to compensate for the sunk costs of establishing the programme in the cancelled countries. In the case of South Sudan, the results from that country were permitted to be made up through progress in other countries. In the case of the drought in East Africa, an allowance was made in the payment decision after the verification team determined that the severity of the drought was unusual and ‘could not have been expected during the period of the WASH Results Programme’, and recommended the full payment be made despite the full outcome-level results not being achieved. Negotiations around these events were time-consuming and required the suppliers, verifiers, and DFID to carry out additional assessments.

Did the use of the PbR modality contribute to DFID market-shaping objectives?

One of DFID’s key aims in using PbR for the WRP was to build the supplier base and to attract new market entrants. The WRP did identify NGOs capable of managing WASH programmes at a scale comparable to the United Nations Children’s Fund, which provides one of the best benchmarks for larger-scale donor-funded WASH programmes. However, given the limited number of NGOs able to demonstrate the ability to manage such large projects, market-shaping was limited. All the successful lead suppliers were organisations that had either significant WASH programmes in the proposed countries, allowing them to better manage financial risks, or were established consortia. No private sector organisation led a successful bid under the WRP, and, overall, the private sector response to this opportunity was more limited than had been anticipated by DFID.

One of the challenges in commissioning PbR projects is the availability of reliable price benchmarks for tender evaluation, which was a problem at the onset of the WRP. As the suppliers worked across a wide range of contexts, it was challenging to assess what were reasonable variations

in price caused by the delivery location, the particular groups targeted, and other qualitative factors.

DISCUSSION

The lessons from the WRP raise some interesting questions about the application of PbR in WASH programmes, and the use of PbR more broadly. The WRP experience suggests that the greatest value of PbR was in focusing supplier attention on specific aspects of programme implementation.

Flexibility in implementation

A key theorised benefit of PbR is that it allows for greater innovation (DFID 2014a, 2014b). This assertion is rooted in the idea that because suppliers are only accountable for results and are not tied to reporting against activity-based workplans, they are able to propose novel approaches and have greater flexibility in implementation. Similar to the findings of Albertson *et al.* (2018b) and the broader emerging evidence base, the WRP experience indicates that this does not occur in practice. PbR tends to promote risk-averse behaviour by suppliers, who stick to tried and tested approaches and locations in which they have greater confidence in their ability to deliver expected results, and so secure payments. It is possible that the strict and pressured time limits for achieving results, the limited experience in the use of PbR and the structure of the contracts reinforced this behaviour. Funders seeking to support the development of genuinely novel and innovative approaches must acknowledge the need for a ‘space to fail’. In such cases, where ‘failure’ is a very real possibility of the financial risks associated with PbR appear likely to become unacceptable for both funder and supplier.

Stronger monitoring and verification

System-based verification reinforced incentives for suppliers to improve their monitoring but potentially came at the cost of securing ‘hard’ independent data. This was a trade-off that was supported by the spot-checking of results by the verifiers to provide confidence that monitoring systems were reliable and robust. The use of spot-checks allowed issues of concern

to be raised and discussed at payment decisions meetings, and in some cases resulted in deferred payments until such time as suppliers could provide reliable data. The alternative of verification through primary independent data collection would have significantly raised costs to the funder and would have reduced the incentives for suppliers to make their own reporting more robust.

The limited inception phase created issues for the efficiency of the monitoring. A longer inception phase would have allowed the more complex elements of verification of systems and in particular outcomes to be addressed. Although the suppliers underestimated the additional costs of improving monitoring systems, for the WRP, in the view of the authors, the benefits from improved quality of data from better monitoring systems and strong technical verification justified the additional costs. However, it is noted that these benefits were largely confined to improvements within the NGO partners, and there was limited impact on government monitoring practices and systems in the countries in which the suppliers were working.

Impact on outputs and outcomes

The WRP experience confirms the findings of [Clist \(2017\)](#) that PbR works best when it is focused on issues of shared concern between supplier and funder and that selecting the right performance measures is critical to the success of a PbR programme. However, linking payment to the early 'results' packages comprised of activities was inefficient to verify and of little benefit to improving programming. This suggests that future programmes using PbR should consider using a mixed model of grant and PbR funding, applying the former to essential inputs and the latter to the delivery of target output results. Identifying strong performance measures at the outcome level proved more difficult than identifying those at the output level, which was partly because outcome measures were not agreed at the outset of the contracts. In future PbR programmes, greater attention should be given to resolving these questions early in design, and before programming starts.

In designing effective PbR WASH programmes, careful consideration is needed regarding the proportion of total funding that uses PbR, and what proportion is linked to output and outcomes. Higher-risk PbR contracts

(those with a large proportion of payment specified on outcomes) may raise the risk profile to unacceptably high levels for suppliers, who may, therefore, choose not to bid on contracts. Conversely, if the PbR element is too small, it may fail to offer real incentives. In terms of longer-term sustainability, PbR may be less suited for programmes focused on system-building, as the indicators that currently exist for measuring systems development are related to processes and activities rather than outcomes. They rely on qualitative evidence and subjective judgement, making them challenging and expensive to verify. Furthermore, unless the PbR contract is with a government, the supplier will have very little direct control over the results.

This suggests that PbR is not suited for use in all WASH programmes; its use should be governed by the specific rationale for using PbR, and in instances where reliable and verifiable outcome indicators are available. Where PbR is applied at both the output and outcome level, as in the WRP, it is important to consider how these are linked. For SAWRP and SWIFT, the outcome targets were defined as proportions of the contracted output-level beneficiaries rather than as a proportion of the actual output-level beneficiaries reached. As such, where output targets were substantially exceeded, as they were in many countries, the outcome targets became substantially less demanding. In future PbR programmes, it would be preferable to base outcome targets on the proportion of the actual output population reached. The experience of the WRP in the face of unexpected events – Ebola in West Africa, Violence in South Sudan, and drought in East Africa – highlighted that PbR afforded no greater benefit in being able to operate in fragile contexts and in each of the cases mentioned that the programme was either suspended or allowances were made in payment decisions. The WRP findings suggest that, overall, PbR is likely to be better suited to non-fragile contexts and if the application of PbR in fragile states is pursued, it is likely that more extensive risk assessment would be required at the design stage.

The WRP only used downside incentives, though there are arguments for the use of upside incentives in PbR programmes from a supplier perspective (it lowers the risk profile while maintaining a focus on results), to motivate suppliers to achieve more results than planned, or if the

purpose of the PbR modality is to incentivise improved quality in implementation. However, upside incentives can pose challenges for funders, who need predictability in multi-year budgets, especially when these funders are government agencies. Donor financial systems are often better at accommodating uncertainty associated with downside incentives. For instance, previous behaviour can be modelled to estimate whether suppliers have a record of under-spend, and this can be used as the basis for budget projections. This permits the reallocation of available funds on a month-by-month basis. Upside incentives may create greater uncertainties for the funder, which are propagated into the future and have an impact on medium-term budget profiles. This tends to have a more significant impact on budget accuracy and planning, and as such while upside incentives may be preferred by suppliers, and offer different incentives to purely downside incentives, they can be challenging to execute for some donors.

Market shaping

Similar to the findings of [Albertson *et al.* \(2018b\)](#), market shaping under the WRP gave somewhat mixed results. New large contracts were established, but the overall number of NGOs realistically able to manage such large, time-sensitive contracts was small. It is unclear whether NGOs willing to bid for such large contracts would do so in a context in which they would be moving into a country without pre-existing programmes. This suggests that non-geographically specific programmes may benefit from a wider market, but programmes focused on individual countries may not. It is noteworthy that no private sector organisation led a successful bid under the WRP, and, overall, the private sector response to this opportunity was more limited than had been anticipated by DFID.

CONCLUSION

The WRP has shown that PbR can be an effective financing instrument for rural basic WASH projects. It has delivered significant results at the output level and shown that these can be sustained for at least 2 years post-implementation at high levels. PbR was effective in promoting investment

in monitoring systems by suppliers and independent verification increased the quality and reliability of the results. As with the previous use of PbR, it is essential that strong performance measures are agreed that address issues that both suppliers and funders care about. The experience under the WRP highlights that the PbR modality did not stimulate innovation and, in fact, promoted risk-averse behaviour by suppliers. The WRP experience also indicates that future WASH programmes using PbR will need to incorporate an inception phase to agree performance measures and processes of verification, and to obtain clarity on outcomes. It is also recommended that future programmes only apply PbR to a portion of total funding reserved for incentivising specific aspects of programming and that essential project inputs are funded through grant mechanisms. There is a need for further evaluations of PbR in practice to establish where, how, and in what ways this modality can be most effectively used, and in the coming years to synthesise the emerging evidence from the PbR programmes that have finished and been evaluated. The question of whether PbR should be used in preference to other modalities remains unanswered and will depend on the context of the programme design. Where there are clear targets and no expectation of innovation, then PbR is well suited. Where objectives are more complex or where innovation is desired, other financing mechanisms may be more appropriate.

ACKNOWLEDGEMENTS

This work draws on the evaluation funded by DFID, primarily under the contract Monitoring, Verification and Evaluation Service Provider for the WASH Results Programme. The views expressed in this paper are solely those of the authors, they do not represent the position of DFID or OPM. The authors would like to acknowledge the wider group of authors responsible for the WRP evaluation, which included: Julia Larkin, Lucrezia Tincani, Jeremy Colin, Sue Cavill, Sarah Javeed, Alice Mango, Faith Muniale, Shona Jenkins, Alex Hurrell, Sophie Witter, Richard Carter, and Timothy Forster.

REFERENCES

- Albertson, K., Fox, C., O'Leary, C., Painter, G., Bailey, K. & Labarbera, J. (2018a) *Introduction: outcome-based payment and the reform of public services*. In: *Payment by Results and Social Impact Bonds: Outcome-Based Payment Systems in the UK and US*. Policy Press, Bristol. doi:10.1332/policypress/9781447340706.003.0001.
- Albertson, K., Fox, C., O'Leary, C., Painter, G., Bailey, K. & Labarbera, J. (2018b) Review of the evidence for outcome-based payment systems. In: *Payment by Results and Social Impact Bonds: Outcome-Based Payment Systems in the UK and US*. Policy Press, Bristol.
- Cabinet Office 2011 *Open Public Services White Paper*. Cabinet Office, London.
- Clist, P. 2017 *Review of Payment by Results in DFID: Establishing the Evidence Base*. DFID, London. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684278/full-report-UEA2-merged.pdf.
- DFID 2013 *Water, Sanitation and Hygiene (WASH) Results Programme: Business Case and Intervention Summary*. Available from: <https://devtracker.dfid.gov.uk/projects/GB-1-203572/documents>.
- DFID 2014a *Sharpening Incentives to Perform: DFID's Strategy for Payment by Results*. DFID, London.
- DFID 2014b *Designing and Delivering Payment by Results Programmes: A DFID Smart Guide*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/352519/Designing-Delivering-PbR-Programmes.pdf.
- DFID 2016 *Water, Sanitation and Hygiene results programme to support scale-up efforts, Annual Review 2016*. Available from: <https://devtracker.dfid.gov.uk/projects/GB-1-203572/documents>.
- DFID 2018 *Water, Sanitation and Hygiene Results Programme to Support Scale-Up Efforts, Annual Review 2018*. Available from: <https://devtracker.dfid.gov.uk/projects/GB-1-203572/documents>.
- ePact 2018 *WASH Results Programme Evaluation – Synthesis Report Vol. 1.1*. Available from: <https://www.gov.uk/dfid-research-outputs/wash-results-programme-evaluation-synthesis-report>.
- Fox, C. & Morris, S. 2019 Evaluating outcome-based payment programmes: challenges for evidence-based policy. *Journal of Economic Policy Reform*. doi:10.1080/17487870.2019.1575217.
- Hood, C. 1991 *A public management for all seasons?* *Public Administration* **69** (1), 3–19. doi:10.1111/padm.1991.69.issue-1.
- National Audit Office 2015 *Outcome-Based Payment Schemes: Government's Use of Payment by Results*. National Audit Office, London.

First received 24 February 2020; accepted in revised form 24 April 2020. Available online 19 May 2020