



Department
for International
Development



DFID's Evaluation Framework for Payment by Results



Contents

Executive Summary	3
1. Introduction	4
2. Expectations for Payment by Results	5
3. Strength of Evidence	10
4. DFID's Priority Evaluation Questions	13
5. Selecting Evaluation Approaches and Methods	15
6. Comparison, Synthesis and Communications	19
7. Conclusions and Next Steps	23

Executive Summary

DFID believes that PBR is an innovative delivery mechanism which can, in the right circumstances, be used to sharpen incentives for suppliers, and partner governments, to perform. Where it works best, PBR can shift the focus of commissioners and suppliers from inputs to outcomes – leading to real, tangible, measurable improvements in the areas which matter most for poverty reduction. DFID is committed to expanding the use of Payment by Results. We believe that the most effective way to do so – and to deliver increasing value for money and development impact for each pound DFID spends – is to build evidence on what works best and in what circumstances.

This Evaluation Framework supports DFID's Strategy for Payment by Results by setting out an agenda for using evaluation to learn about what works and ensuring value for money. This Framework is intended for DFID and wider communities interested in understanding a broader range of aid instruments referred to as 'Payment by Results' (PBR).

This Framework does not replace any existing policies, strategies, quality standards or guidance for the evaluation of PBR programmes. Instead, it complements existing information by identifying the evidence gaps and learning opportunities that are specific to Payment by Results.

The Framework first sets out the context for PBR instruments. It then uses existing expectations and evidence as the basis for selecting Priority Evaluation Questions that can be answered within and across programmes. These Priority Evaluation Questions are intended to be additional to other context specific evaluation questions.

DFID recognises that new and innovative PBR instruments are being developed in contexts and for sectors that are rapidly changing. To maximise learning without stifling innovation this Framework intends to make it easier to design context relevant evaluations that also help DFID learn across programmes. Deciding how to answer Priority Questions in a given context should be done as appropriate and with technical evaluation support.

1. Introduction

1.1 This Evaluation Framework provides guidance on why and how to evaluate Payment by Results programmes. The purpose of this Evaluation Framework is to encourage generating evidence about Payment by Results instruments that is useful for designing future interventions.

1.2 DFID categorises interventions as Payment by Results when payments are made for pre-agreed results, rather than inputs or milestones (See Table 1 in Section 2). The Evaluation Framework is particularly concerned with payments made for agreed outcomes, since this is the most innovative form of Payment by Results, but covers payments made for agreed outputs and processes too.

1.3 In June 2014 DFID published *Sharpening incentives to perform: DFID's Strategy for Payment by Results*¹. Payment by Results is part of wider HMG reform². DFID's Strategy sets the objectives of building the evidence base for PbR and building capabilities to do it well, with the aim of ensuring value for money in development budgets by paying on delivery of results, whenever appropriate. To learn what types of Payment by Results work best, DFID is committed to conducting rigorous, independent evaluations suitable for accumulating knowledge across interventions.

1.4 Evidence supporting Payment by Results is at an early stage, and the strength of evidence varies between sectors and types (see page 6). Because DFID is committed to learning what instruments are best suited for development, evaluations are needed that identify and measure the effects of Payment by Results (instrument effect). This will be achieved by:

- i. Evaluating Payment by Results interventions to identify and test to what extent expected outcomes are caused by the payment approach and how;
- ii. Accumulating evidence across interventions to understand the likelihood (e.g. probability) of Payment by Results having desired outcomes in different contexts;

¹ DFID. 2014. *Sharpening incentives to perform: DFID's Strategy for Payment by Results*.

<https://www.gov.uk/government/publications/dfids-strategy-for-payment-by-results-sharpening-incentives-to-perform>.

² Payment by Results is strongly referenced in the Cabinet Office's [Open Public Services White Paper](#), and [Open Public Services 2012 Update](#), which set out the Government's priorities for civil service reform.

- iii. Comparing Payment by Results with other available aid instruments to establish appropriateness and value for money in different development interventions.

1.5 The following sections examine expectations and evidence as the basis for selecting Priority Evaluation Questions and guidance on how to answer them.

2. Expectations for Payment by Results

Context and Global Expectations for Payment by Results

2.1 The concept of Payment by Results is growing increasingly popular in development cooperation globally under many different names, for example ‘cash on delivery’, ‘payment for results’, and many others. Definitions and ways of implementing Payment by Results differ greatly in accordance with underlying expectations and theories of change for how the aid instrument will affect development partnerships.

Table 2.1 Example Theories of Change for Payment by Results³

Goals	Theories of Change
Accountability	Increasing visibility of results → <u>Increased assurance</u> to donors and beneficiaries
Incentives	Payment for results → <u>Aligning incentives</u> to increase effort of implementers to achieve results
Flexibility / efficiency	Focusing on results → Allow space for adaptive use of inputs and processes to <u>efficiently</u> achieve results
Innovation	Targeting difficult-to-achieve results → Allow space for suppliers to pilot, monitor and evaluate different <u>innovative approaches</u> , scale up successes and scale down failures

2.2. Table 2.1 is not an exhaustive list of potential theories of change that could motivate use of payment by results as a modality and the theories of change are not mutually exclusive. Each example that it highlights will potentially require different approaches to programme design, delivery, monitoring and management to yield the different benefits sought from using payment by results.

³ Theories of Change examples are based on presentations and discussion at the Payment By Results: Theory to Evidence workshop, 21 November 2014, London

2.3 Given this diversity, DFID encourages anyone engaged in designing an innovative programme which uses Payment by Results to reflect on the theory of change of Payment by Results, as a subset of the broader theory of change of the intervention. Carefully thinking through the logical steps by which a Payment by Results mechanism will lead to, or improve, outcomes, in the particular context of the programme, will improve programme design and increase the scope for effective evaluation.

How DFID understands Payment by Results

2.4 While definitions of Payment by Results differ across donors and international organisations, DFID categorises interventions as Payment by Results if payments are made after the achievement of pre-agreed results. As outlined in the 2014 Strategy, DFID distinguishes between three types of Payment by Results, varying with implementers (see Table 1).

Table 2.2: Overview of different types of Payment by Results

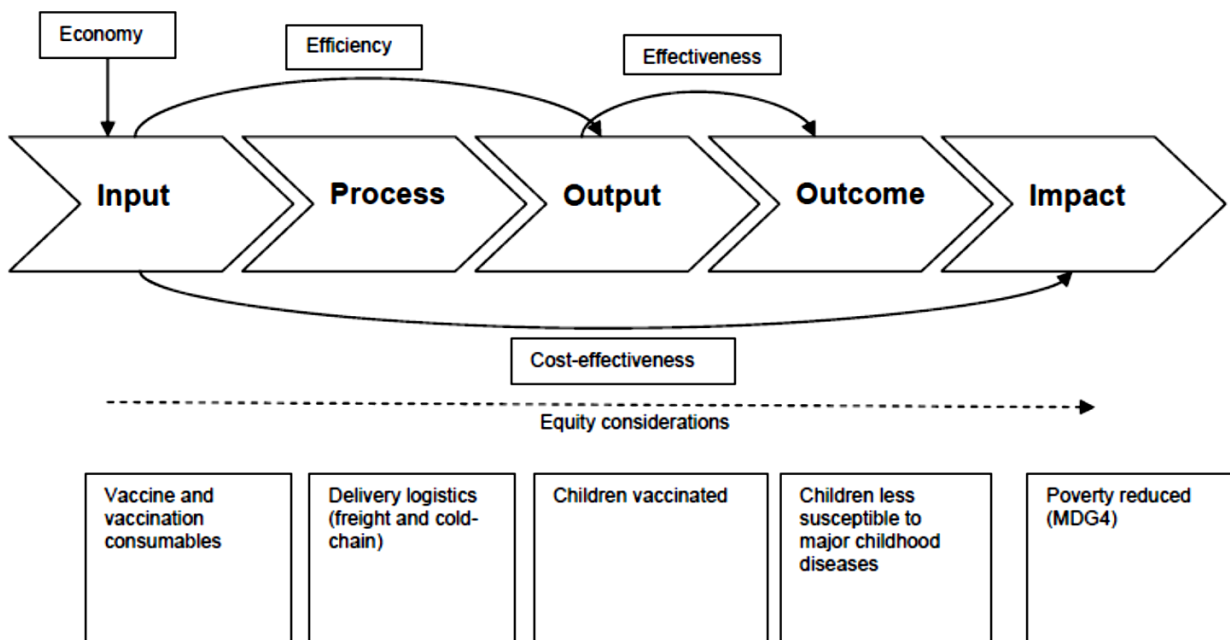
	Results Based Finance	Results Based Aid	Development Impact Bonds
Agent models	<pre> graph TD OP1[Outcome Payer (e.g. DFID)] --> SP1[Service Provider] SP1 --> TP1[Target Population] VA1[Validating agency] --> OP1 VA1 --> TP1 </pre>	<pre> graph TD OP2[Outcome Payer (e.g. DFID)] --> RG[Recipient government] RG --> TP2[Target Population] VA2[Validating agency] --> OP2 VA2 --> TP2 </pre>	<pre> graph TD I[Investor] --> Int[Intermediary] Int --> SP[Service Provider] SP --> TP3[Target Population] OP3[Outcome Payer (e.g. DFID)] --> Int OP3 --> TP3 VA3[Validating agency] --> OP3 VA3 --> TP3 </pre>
Risk transfer to	Service provider	Partner government	Investor

2.5 The theory and expectations around all of these types of Payment by Results instruments rest on two essential assumptions⁴:

1. There is a risk transfer as payment depends on an output or an outcome, not action.
2. Payment is contingent on (independently) verified results⁵

⁴ The theory base and expectations in this chapter mainly draw on the following DFID-funded study: Clist, P.; Verschoor, A. 2014. The Conceptual Basis of Payment by Results in International Development. Norwich: University of East Anglia.

Figure 2.3: DFID understanding of economy, efficiency and effectiveness⁶



2.6 The theory behind Payment by Results suggests these approaches may have a number of different benefits delivered in different circumstances:

- Improved efficiency and effectiveness of delivery from the incentives in a Payment by Results contract;
- Increased potential for innovation and flexibility in delivery through not specifying inputs;
- Increased transparency and accountability for results and a strong focus on performance;
- Increase in value for money of interventions.

2.7 Understanding when, why and how these expected benefits are likely to materialise requires looking at the theory and assumptions behind RBA, RBF and DIBs.

⁵ DFID's Payment by Results strategy defines results as outputs or outcomes, or processes leading to an improvement in performance

⁶ The term 'impact' in a results chain is often different from the focus of an 'Impact Evaluation', which may measure the effect of an intervention on 'outcomes' or 'impacts'.

Potential benefits and risks of RBA and RBF initiatives

2.8 Results-based Aid and Results-based Finance programmes can be distinguished only insofar as payments are made to different implementers (governments for RBA, and other kinds of service provider for RBF – for example private sector firms or NGOs).

2.9 The theory behind Payment by Results builds on economic theory which looks at the relationship between donors and implementers. From this perspective, RBA and RBF are expected to create stronger incentives for implementers to undertake desired actions, but at the same time impose greater risk upon them. The trade-off for the donor when considering a results-based contract is whether the positive gains outweigh the extra costs in the form of a risk premium paid to the implementer.⁷ Other economic models build on the above-mentioned assumptions, but focus on the performance measure itself and analyse which rewards induce desired actions. While economic theory is useful when analysing the contractual arrangements of Payment by Results programmes, political economy approaches help in understanding practical constraints, such as the difficulty of withholding aid.

2.10 However, the extent that expected benefits are realised depends on at least 12 principles⁸, of which the following are particularly problematic and need supporting evidence from practice:

- **Quality of the performance measure:** The quality of the performance measure is a key determinant of the potential benefit of RBA and RBF programmes. In order for Payment by Results programmes to be effective, the performance measure needs to be correlated with the underlying variable of interest *ex ante* and *ex post*.
- **Alignment:** RBA or RBF contracts can be more beneficial than regular contracts where there is incomplete alignment between principals and agents. Otherwise, the agent would undertake the measure independently of the agent, and the expected efficiency gains of Payment by Results contracts are unlikely to materialise.

⁷ Clist, P.; Verschoor, A. 2014. *The Conceptual Basis of Payment by Results in International Development*. Norwich: University of East Anglia.

⁸ Clist, P.; Dercon, S. 2014. *12 Principles for Payment by Results in International Development*. London: Department for International Development.

- **Observability of effort:** If the agent's costly actions can be directly observed RBA or RBF programmes might not be beneficial over regular interventions, as the principal can then contract upon these inputs directly.
- **Control:** RBA and RBF are more likely to be beneficial where agents have a larger degree of control over the outcomes (a high signal-to-noise ratio), as otherwise a large prize would be needed to incentivise effort.
- **Risk aversion and risk transfer:** Smaller agents are likely to be more risk averse, as they face proportionally higher losses in case of non-reward. This risk preference of agents will affect the cost to the principal, and it may also reduce the size of the pool from which an agent may be drawn. However, at the same time risk transfer from a principal to an agent is never complete as the principal can still suffer reputational or political risk.
- **Distortion and Gaming:** In Payment by Results interventions the gap between the measured results upon which payments are made and the true efforts of an agent, which are not fully observable, could lead to gaming. However, as Clist/Verschoor (2014) argue, distortion is more likely to be an issue than gaming, as the use of a performance measure may lead to an efficiency loss by over-rewarding actions related to the measured result and under-rewarding actions on aspects of the result that are not measured, possibly distorting incentives. Distortion is a further concern in RBA and RBF, as we may have evidence of an improvement but cannot necessarily be confident about whether the underlying goal has been positively or negatively affected.
- **Verification cost:** Relative to other forms of aid, verification costs of Payment by Results programmes are likely to be higher. This may, however, be offset by other potential efficiency gains, such as the reduction in transaction costs and management time.
- **Macro-level risks and limitations:** In development interventions the threat of non-payment is not always credible. If the donor is unable to withhold payments to the agent, RBA or RBF is unlikely to be more effective than other instruments.

Potential Benefits and Risks of Development Impact Bonds (DIBs)

2.11 While DIBs are similar to RBA and RBF insofar as disbursement of funds is based on pre-agreed results, the additional agents and contractual arrangements (see Table 1) lead to a distinct incentive and risk structure. In DIBs, an investor (not the donor) provides upfront financing to an intermediary. The intermediary uses this money to pay service providers to affect some target population. The outcome payer (donor or principal) will then disburse funds to the investor on the

basis of measured success as reported by the validating agency. It is worth noting that the possible range of different contractual arrangements for DIBs is vast.

Due to this different model, some additional benefits and risks can be expected with DIBs:

- **Risk:** One of the arguments for DIBs is that they have the potential to provide a different allocation of risk amongst the different parties compared to RBA/RBF. For example, social impact investors may be better able to accept some forms of reputational risk associated with failure to deliver than could government donors or NGO suppliers.
- **Market effect:** Another potential benefit unique to DIBs is their potential for catalysing a growing social investment market. In theory, once a certain number of DIBs have been successfully rolled out it is likely that more interested investors, intermediaries and outcome payers will be willing to engage with similar bonds, and a DIB market can emerge.

2.12 Contractual complexity: Due to the larger number of agents (see Table 1) and many possible arrangements between these agents DIB contracts have the potential to be significantly more complex than RBF and RBA programmes. While higher initial set-up costs may reduce over time, DIBs face a higher risk of foregoing the expected efficiency or effectiveness gains outlined above.

3. Strength of Evidence

Evidence base behind Payment by Results

3.1 The evidence base supporting Payment by Results is emerging from many global initiatives. Because of the greater past activity in health and education, the majority of existing evidence on Payment by Results comes from RBF programmes in these two sectors.

3.2 A DFID-commissioned study, Evaluation of Payment by Results (2013)⁹, found at that time that most Payment by Results evaluations could benefit from an increased focus on impact and value

⁹ Perrin, B. 2013. Evaluation of Payment by Results (PBR): Current Approaches, Future Needs. London: Department for International Development

for money¹⁰. Many Payment by Results evaluations had not identified the additional benefits of the PBR mechanisms (instrument effect), which can also be complicated by measurement difficulties.¹¹

Table 3.1: Selection of evidence for different types of Payment by Results

	Results Based Finance	Results Based Aid	Development Impact Bonds
Strength of Evidence	Medium (mainly in health)	Emerging	Untested
DFID Portfolio (example)	Uganda: RBF health programme; DFID-funded Health Results Innovation Trust Fund (HRITF)	Rwanda: RBA programme in education	Uganda: Sleeping sickness DIB (inception phase)
Examples from other donors	GPOBA (World Bank & other donors) - Service delivery is contracted out to third party (usually private sector) that receives a subsidy to complement or replace user fees	World Bank Programme for Results (PforR) - funds for specific expenditure programmes of partner countries; “Disbursement Linked Indicators” (DLIs) and verifying agency;	No other existing DIBs; few existing examples of similar Social Impact Bonds in the UK and the USA (e.g. Peterborough SIB)
	Multiple RBF health programmes by different donors (EU, USAID, DANIDA and various NGOs)	European Commission MDG Contracts - Payment made to government in return for commitment to good governance and progress in poverty reduction; EU More for More Approach in EU neighbourhood policies	
	GFATM (Hybrid) - Funding for years 3 to 5 dependent on overall performance achieved during first two years of grant implementation		

3.4 To better focus evaluation it is helpful to reflect on the evidence for each type of Payment by Results:

- **RBF:** The existing evidence base in RBF comes largely from health programmes. A number of literature reviews¹² show RBF health programmes tend to focus on easier to measure vertical outcomes (e.g. number of vaccinations), than harder to measure outcomes like health system improvements. One review finds that contracting-out health services may increase access and

¹⁰ Ibidem; and Pereira, J.; Villota, C. 2013. Hitting the target? Evaluating the effectiveness of results-based approaches to aid. Brussels: EURODAD.

¹¹ Pearson, M. 2011. Results based aid and results based financing: What are they? Have they delivered results? London: HLSP.

¹² Grittner, A. 2013. Results-based Financing. Evidence from performance-based financing in the health sector. Bonn: Deutsches Institut fuer Entwicklungspolitik; Norwegian Knowledge Centre for Health Services. 2008. An overview of research on the effects of results-based financing. Oslo: NKCHS;

use¹³, and another that RBF can potentially improve the quality of services.¹⁴ Further evidence is needed to understand how RBF leads to the better performance and efficiency that was observed.¹⁵

- **RBA:** RBA is a relatively new instrument and the existing evidence draws largely on monitoring data. However, a rigorous study of an RBA initiative in Indonesian villages does find positive effects on shorter-term health and education indicators¹⁶. The weakening of effects over time suggests indicators and the timeline for collecting data is important. The study does not find evidence of gaming, and finds positive effects on budget allocation.

RBA decentralisation programmes in Ghana and Tanzania found “positive results” in terms of financial governance and the engagement between local governments and citizens.¹⁷ A mid-term evaluation of a DFID’s Rwandan education programme found that the pilot was highly relevant to target groups and induced a high level of government ownership, but more time is needed to observe the impact on education.¹⁸

- **DIBs:** Early evidence for DIBs comes from Social Impact Bonds (SIBs), which differ because they are not in developing country contexts. The use of SIBs has been expanding and 15 have been documented in the United Kingdom.¹⁹ Much existing evidence comes from an intervention at Peterborough prison, which aimed to reduce reconviction rates. A first evaluation of Peterborough finds positive effects²⁰ and confirms that developing indicators is time-consuming

¹³ Perrin, B. 2013. Evaluation of Payment by Results (PBR): Current Approaches, Future Needs. London: Department for International Development

¹⁴ Gorter AC et al. 2013. Evidence Review: Results-Based Financing of Maternal and Newborn Health Care in Low- and Lower-middle-Income Countries. Study commissioned by the German Federal Ministry for Economic Cooperation and Development.

¹⁵ Pearson, M. et al. 2010. Review of major Results Based Aid (RBA) and Results Based Financing (RBF) Schemes. London: Department for International Development

¹⁶ Olken, B.; Onishi, J.; Wong, S. 2014. Should Aid Reward Performance? Evidence From a Field Experiment on Health and Education in Indonesia. *American Economic Journal: Applied Economics* 6 (4).

¹⁷ Janus, H. 2014. Real Innovation or Second-Best Solution? First experiences from results-based aid for fiscal decentralisation in Ghana and Tanzania. Bonn: Deutsches Institut fuer Entwicklungspolitik.

¹⁸ Upper Quartile. 2014. Evaluation of Results Based Aid in Rwandan Education - 2013 Evaluation Report. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/312006/Rwanda-education-results-based-aid-evaluation

¹⁹ Clist, P.; Drew, R. 2014. Evaluating Development Impact Bond. A Study for DFID. London: Department for International Development.

²⁰ Disley, E.; Rubin, J. 2014. Phase 2 report from the payment by results Social Impact Bond pilot at HMP Peterborough. London: MoJ.

and difficult.²¹ There are outstanding questions about Peterborough and the external validity of findings.²²

4. DFID’s Priority Evaluation Questions

4.1 DFID is committed to answering the following broad Priority Evaluation Questions:

1. What payment by results instruments work best in different circumstances?
2. When and how do payment by results incentives work in practice?
3. What is the value for money of different types of payment by results instruments?

4.2 Each DFID-funded evaluation is guided by OECD DAC evaluation criteria²³ and programme evaluation questions. In addition to programme-specific questions, these Priority Evaluation Questions should be considered when evaluating Payment by Results programmes. However, not every Payment by Results programme aims at realising all of the potential benefits stemming from the instrument. In some cases Payment by Results will be used to improve delivery, while in others it encourages innovation. The purpose of using a Payment by Results instrument and the level at which results will be targeted (processes, outputs, outcomes) should be clearly outlined at the design stage, and will influence the Theory of Change as well as the appropriate evaluation questions:

Expectations	Evaluation Question(s)	Comment
<i>Interventions uphold DFID policy and produce intended outcomes</i>	To what extent and for whom did the intervention achieve intended outcomes? Are the outcomes indicators still relevant? Were there any unintended negative outcomes to beneficiaries – for example to marginalised or harder-to-reach groups? Did it uphold policy (for instance gender, environment)?	
<i>PbR mechanism alter incentives of suppliers and/or recipients</i>	Did PbR incentives impact (positively or negatively) on behaviour of the implementer? Did incentives have the anticipated effect on processes, outputs or outcomes? If not, was this due to insufficient incentive, inappropriate design, or factors out-with the control of the procurer, supplier or recipient?	<i>Theory suggests distortion and gaming are particular concerns where achievement of proxy indicators are used to trigger payment, rather than direct measurement of outcomes</i>

²¹ Disley, E. et al. 2011. Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough. London: MoJ.

²² Clist, P.; Drew, R. 2014. Evaluating Development Impact Bond. A Study for DFID. London: Department for International Development.

²³ OECD DAC evaluation criteria include: relevance, effectiveness, efficiency, impact, and sustainability.

Expectations	Evaluation Question(s)	Comment
	<p>Were there unanticipated (positive or negative) effects or consequences of the instruments on the implementer, and / or on the beneficiary?</p> <p>Was there evidence of distortion and/or gaming?</p>	
<p><i>Improved economy, efficiency or effectiveness of delivery</i></p>	<p>Did the PbR mechanism allow the supplier to achieve economies (lower cost procurement of inputs) beyond what were expected?</p> <p>Are these lower prices replicable, and if so useable as benchmarks for other DFID procurement negotiations?</p>	<p><i>PbR may allow suppliers greater flexibility in inputs than more traditional contracting – potentially allowing discovery of economies.</i></p>
	<p>Did the PbR mechanism lead to improvements in efficiency – in converting inputs to outputs?</p> <p>Was there scope to capture the benefits of any efficiency improvements in the pricing mechanism, e.g. through gain-share arrangements?</p> <p>How do the unit costs per output for the supplier compare to alternative procurement mechanisms delivering similar outputs?</p> <p>If costs were low, are they replicable, and if so useable as benchmarks for other DFID procurement negotiations?</p>	<p><i>Particularly important for PbR programmes where disbursement is measured on the achievement of outputs, such as years' schooling</i></p>
	<p>Did the PbR mechanism lead to improvements in effectiveness – in converting outputs to outcomes?</p> <p>Are these improvements replicable?</p> <p>How does value for money²⁴ compare to other available aid instruments?</p>	<p><i>Particularly important for PbR programmes where disbursement is measured on the achievement of outcomes, such as literacy</i></p>
	<p>Did the PbR mechanism lead to greater costs to the donor, supplier or partner government, or beneficiary?</p> <p>Could these additional costs be reduced, without eroding benefits, in future interventions?</p> <p>Were additional costs justified by additional benefits, relative to alternative modalities?</p>	
<p><i>Increased flexibility in delivery</i></p>	<p>Did the PbR mechanism allow the service provider or recipient government more (or less) flexibility to undertake adaptive programme management?</p> <p>Did flexibility in the use of inputs and processes lead to more effective delivery of outputs or outcomes?</p> <p>Was there scope to capture the benefits of innovation in the pricing mechanism, e.g. through gain-share arrangements?</p>	<p><i>This may be important where programme designers and suppliers face uncertainty at planning stage over which of a suite of potential interventions will prove most effective in different circumstances during delivery.</i></p>
<p><i>Increased innovation in delivery</i></p>	<p>Did the PbR mechanism increase (or decrease) the ability of suppliers or partner governments to pilot, monitor and evaluate different <u>innovative approaches</u>, scale up successes and scale down failures?</p> <p>Were innovations successful?</p>	<p><i>This may be particularly important where, for example, neither donor nor supplier or partner government are clear, at planning stage, how best to translate outputs into</i></p>

²⁴ Value for money in DFID's programme means maximising the impact of each pound spent to improve poor people's lives.

Expectations	Evaluation Question(s)	Comment
	Are innovations replicable in future programming?	<i>outcomes.</i>
<i>Increased transparency and accountability</i>	Was there a shift in focus to results and away from inputs? Has the PBR mechanism improved beneficiary feedback, ownership and advocacy? Has the PBR mechanism strengthened the accountability of suppliers and recipient governments to donors?	
<i>DFID and partners have sufficient capacity and time</i>	How did donors and funders ²⁵ support the design and implementation of the intervention? Did donors focus on results instead of inputs, and if so on which type of results (processes, output, outcomes)? Are the results indicators correlated with the underlying variable of interest ex-ante and ex-post intervention? How did the effort required to design and implement the intervention compare with other aid instruments? What was the additional burden or benefit for DFID and for the implementer regarding contract design, and are there varying effects of PBR contracting on different types of implementing organisations? Where there is competitive bidding for PbR contracts, has there been a diverse range of providers competing?	
<i>Stakeholders have the capacity and resources to cooperate in ways which maximises their comparative advantage</i>	How did stakeholders respond to paying for results? To what extent did the decision to use PbR affect the market of suppliers? What roles do stakeholders play in the intervention? To what extent did these roles make maximum use of their comparative advantages?	
<i>Sustainability</i>	How sustainable are the outputs, outcomes and impacts delivered by the programme? How did the sustainability of processes, outputs and outcomes produced compare with other aid instruments?	

4.3 DFID does not expect any one evaluation to answer all of the above evaluation questions. Each Payment by Results initiative will need to first identify the purpose of the evaluation, the theory of change for the selected Payment by Results instruments, and then identify the evaluation questions which can and should be answered.

4.4 In accordance with OECD DAC Evaluation Principles and Evaluation Quality Standards for Development Evaluation, when selecting Payment by Results evaluation questions it is important to consider the needs of relevant stakeholders and intended programme beneficiaries, as they may have different perspectives and evaluations needs.

²⁵ In Development Impact Bonds the funder pays implementers before reimbursement from the donor upon achievement of results.

5. Selecting Evaluation Approaches and Methods

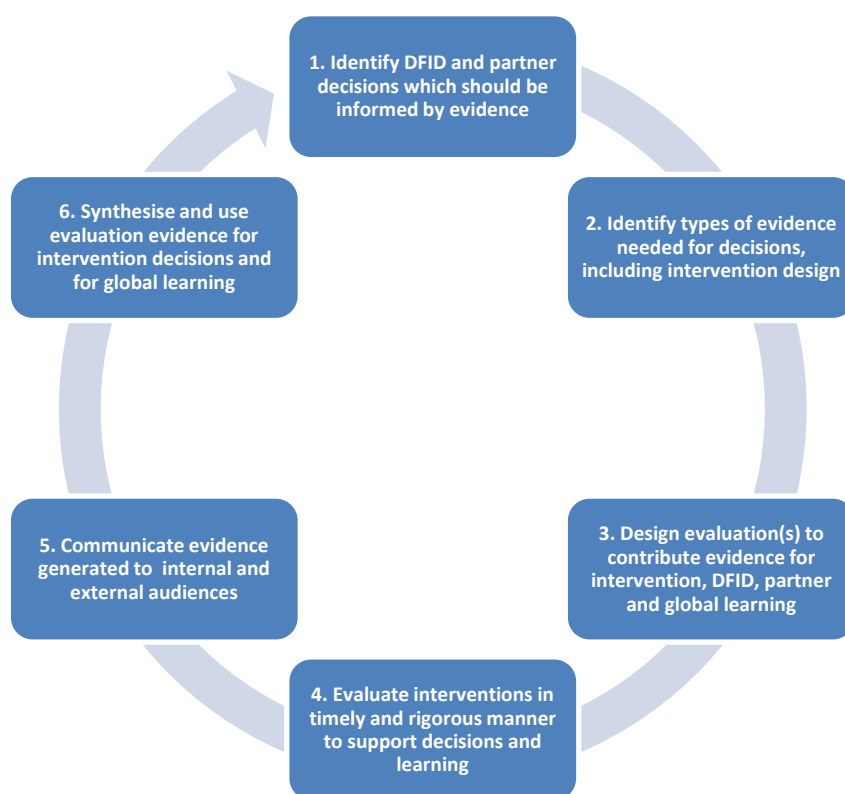
5.1 Each evaluation should be designed to meet the needs of the programme, partners and intended beneficiaries. DFID promotes participation throughout each evaluation cycle.

5.2 Evaluations of Payment by Results programmes should fulfil at least two main purposes:

- i. Generating evidence for programme learning and accountability;
- ii. Generating evidence for understanding the selected Payment by Results instrument(s).

5.3 DFID recognises that effective development programming requires learning how best to end poverty in different contexts. Evaluation is promoted as a tool for learning how to achieve impact and improve interventions, including Payment by Results.

Figure 5.1: Learning focused evaluation cycle

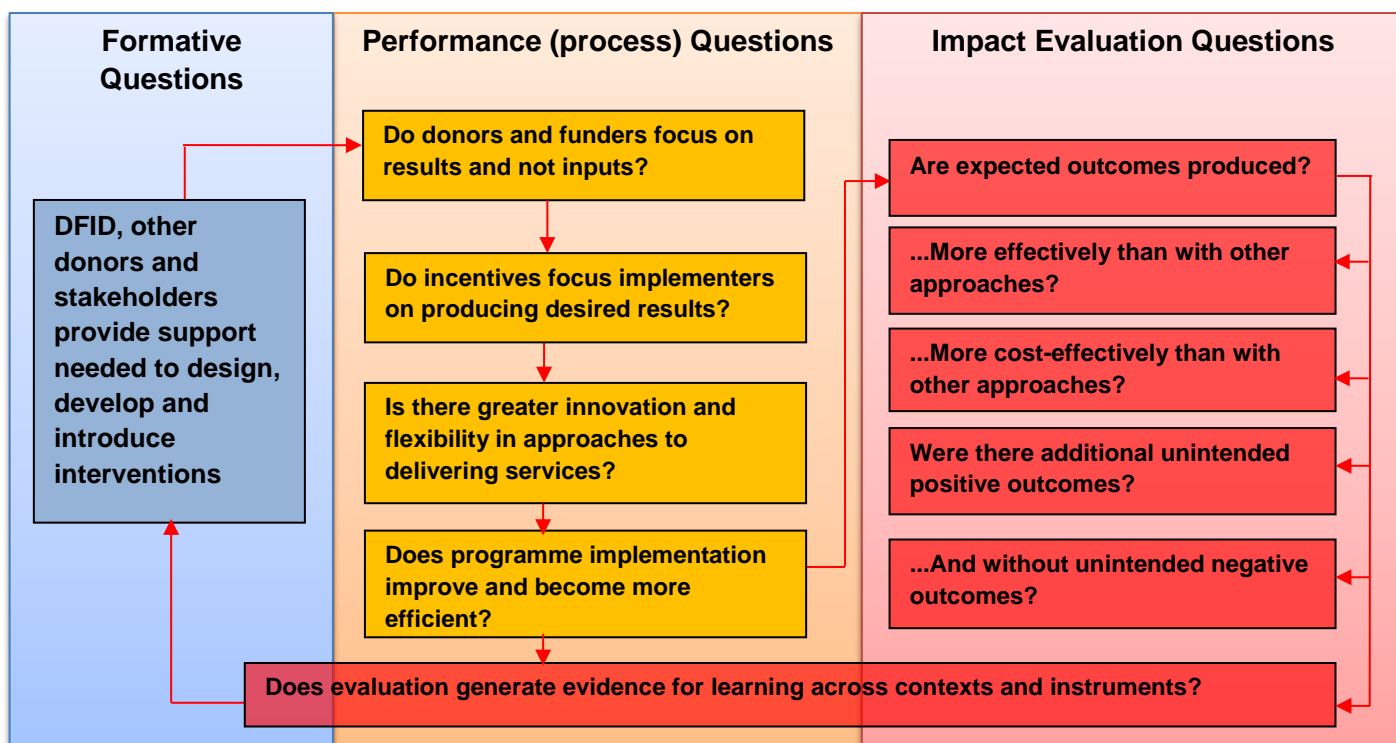


5.4 Each evaluation requires selecting the best approach, methods and tools for answering questions in a way that meets the intended use, purpose and audience. DFID does not subscribe to any one approach and encourages each evaluation to consider the purpose, questions, and programme attributes when selecting an approach, method or tool.

5.5 DFID also requires that each intervention and evaluation be designed in accordance with Paris Declaration principles and takes care to avoid duplication in data collection and evaluation when possible. It is important to consider that Payment by Results has additional costs associated with collecting verification data, and how this data can be incorporated into any evaluation, and the extent to which data can be used to measure the effect on outcomes.

5.6 Each Payment by Results evaluation will use a combination of approaches for answering performance (process) and impact questions: 1) Performance Questions: Has the Payment by Results mechanism improved implementation? 2) Impact Questions: Does the intervention have intended and/or unintended effects on outcomes?

Figure 5.2: Illustration of how Payment by Results questions link to interventions²⁶



²⁶ Impact Evaluations are used to establish the effect of an intervention on intended and unintended long-term or short-term outcomes. As such, the term 'impact' refers to measured effect, which may or may not be the same long-term outcomes listed as 'impacts' in a results framework or theory of change.

5.7 Evaluations can be used before, during and after an intervention. Formative evaluation is used to test what is relevant and most likely to work before implementation at scale. Performance (process) evaluation is focused on understanding implementation. Impact evaluation is focused on the intended and unintended effects of the intervention on measured outcomes, which can be short-term or long-term, intended or unintended.

5.8 When designing an evaluation is it important to consider the purpose of the evaluation and opportunities for stakeholder participation. Payment by results is new for DFID and many of its partners. In cases such as Results Based Aid, where the government partner may have an interest in building its own capacity and understanding of Payment by Results, participatory approaches to evaluation may be most appropriate.

5.9 Different evaluation approaches are used to answer performance and impact questions. An implementation process is evaluated using some form of performance evaluation. Impact questions are answered using experimental, quasi-experimental or non-experimental²⁷ impact evaluation approaches. DFID encourages each Payment by Results evaluation to make best use of both performance and impact evaluation approaches.

5.10 When evaluating Payment by Results it is important to capture performance and impact in ways that inform learning. Ensuring evidence generated in one context has validity in another (external validity) requires capturing outcomes, how they are achieved (i.e. causal mechanisms), and the interaction with the context in which they are produced.

5.11 Mixing methods helps to identify causal mechanisms and can improve understandings of the instrument effects caused by Payment by Results. Many variables and underlying factors that explain the nature and strength of an instrument's effect will be unknown at the beginning of a Payment by Results intervention, and therefore unlikely to be captured by pre-selected data for monitoring and evaluation. Rigorous qualitative evaluation before, during and after an intervention can provide evidence and clues that improve understandings of 'if' and 'how' Payment by Results achieves measured or unmeasured, intended or unintended effects. In turn, identifying clues that explain how instrument effects are achieved can improve quantitative models and the estimation of effects.

²⁷ Non-experimental impact evaluation refers to quantitative and qualitative analytical techniques that do not establish or estimate a counterfactual to measure the effect of interventions. Whereas quasi-experimental approaches use techniques like matching and difference-in-difference to estimate a counterfactual, non-experimental impact evaluations may use regression analysis (e.g. ordinary least squares, multi-level or structural modelling, etc.), process tracing, social network analysis and many other techniques that may not rely on a control or counterfactual.

5.12 DFID recognises that many programmes are not suited to experimental or quasi-experimental impact evaluations. Randomised controlled trials (RCTs) should only be used when the conditions for a successful trial are met, and should be considered before programme implementation (e.g. during design). DFID also promotes quasi- or non-experimental impact evaluations as viable alternatives to RCTs. It should be noted that even where RCTs are appropriate, they may shed little light on the degree to which Payment by Results mechanisms contributed to or hindered the achievement of outcomes. Theory-based performance or process evaluations could be usefully deployed alongside RCTs to understand the chain of causation which led to the achievement (or not) of outcomes.

6. Comparison, Synthesis and Communications

6.1 In addition to evaluating the effect of aid instruments on individual interventions, learning about Payment by Results requires accumulating evidence across interventions on a continuous basis. Accumulating evidence about the intended and unintended outcomes of different types of Payment by Results instrument will support learning by: 1) Enabling lessons learned to inform decisions made within and across interventions; and 2) Enabling synthesis and comparison of what works across contexts and selected aid instruments.

Comparison across interventions and instruments

6.2 Each evaluation provides evidence that can be used for different forms of meta-analysis and comparison. Comparative analysis can be used to identify what causal mechanisms and related factors are linked to Payment by Results instruments above and beyond the many intervention and contextual factors which also effect performance and impact. Comparisons should be considered at three levels:

- Level 1: Within intervention comparison between the performance and impact achieved with and without Payment by Results;
- Level 2: Between interventions that are most similar (e.g. same Payment by Results instrument and sectors in different contexts) or most different (e.g. Payment by Results compared to non-Payment by Results in same sector and context) in relation to the instruments used;

- Level 3: Between bodies of evidence for different types of Payment by Results (e.g. results-based financing versus development impact bonds in education in Africa).

6.3 As DFID and its partners generate evidence for Payment by Results instruments it will be important to synthesise evidence across evaluations to learn the likelihood (probability) that a given instrument is causing performance and impact outcomes.

Within interventions the identification of causal mechanisms linking instruments to outcomes (e.g. risk transfer) can be strengthened by varying both the use of a Payment by Results instrument (i.e. varying the treatment) and other plausible mechanisms (e.g. implementers awarded contracts or size of incentives).

6.4 Comparisons provide important lessons for updating knowledge and assumptions about the Payment by Results instruments. As a body of evidence is generated it will become easier to identify and measure the instrument effect of different types of Payment by Results in different contexts. Better evidence can also improve the selection of Payment by Results instruments during the design of future interventions.

Evaluation synthesis and knowledge accumulation

6.5 Generating evidence that is ready for synthesis and accumulating knowledge across interventions requires evaluating in a way that identifies and tests how change is accomplished. Each type of Payment by Results is intended to alter incentives and promote behavioural changes that will ultimately improve intervention delivery and value for money. Understanding if observed changes are caused by Payment by Results (e.g. is it because of risk transfer?) requires identifying the underlying causal relationship (i.e. mechanism) and testing its effect. Using different approaches and methods is important for understand both 'if' and 'how' interventions result in observed changes.

6.5 Testing the same Payment by Results instruments in different contexts (e.g. replication) enables learning what works across interventions and over time. However, before considering replication it is important to reflect on the following: 1) Is there strong evidence that the instrument is having an effect (e.g. causality) on outcomes? 2) Have the conditions needed (e.g. causal mechanisms, population characteristics, etc.) to achieve an effect been identified and explained? 3) Can the same conditions be identified or created in another context to justify replication?

6.6 It is important that DFID teams reflect on the current state of knowledge for their sector and selected type of Payment by Results before designing an intervention and

accompanying evaluation. To learn about Payment by Results across different interventions and aid instruments it is important that each individual evaluation answers the Priority Evaluation Questions in a way that upholds internal and external validity. As DFID evaluates and learns from individual interventions, knowledge about Payment by Results and underlying causal mechanisms will increase.

6.7 As evaluation evidence is generated it should be synthesised with comparable interventions and evaluations. Levels of comparability between interventions will largely be determined by whether or not the same Payment by Results mechanism can be identified and tested, and whether information is collected about the context under which any effects are measured.

6.8 DFID supports different types of synthesis depending on timing, purpose and wider availability of evidence. Each Payment by Results will benefit from the ongoing synthesis and analysis of data to inform implementation decisions. When evidence is generated from several evaluations and interventions to answer the same question it can be synthesised to examine trends and patterns that are useful for DFID and partners beyond individual evaluations. As many evaluations and similar research studies are completed over time it will be possible to conduct a 'systematic review' to assess the state of knowledge for a given development question.

6.9 DFID will continue to support evaluation and learning from Payment by Results in line with the DFID Payment by Results and Evaluation strategies. Central support for designing and quality assuring all DFID evaluations is provided by the Evaluation Department. Going forwards DFID will actively search for ways to ensure lessons learned from Payment by Results evaluations improve development interventions by increasing opportunities for evidence-based decisions.

Communicating and disseminating evidence

6.9 DFID is committed to ensuring evidence generated through evaluation informs learning for: 1) interventions; 2) programmes; 3) DFID; and 4) wider international development communities. Making evaluation useful for each of these different communities means understanding planning for communication and dissemination at the start of an evaluation cycle.

6.10 Communications is an important part of preparing, designing, implementing and using evaluations. For each Payment by Results intervention DFID encourages staff to consult

the relevant front-line Evaluation Adviser or headquarters at the beginning of the evaluation cycle. During each subsequent step in the evaluation cycle all stakeholders should be engaged in accordance with agreed roles, responsibilities and ethics considerations.

6.11 Dissemination begins at the start of an evaluation. DFID encourages every team to think about relevant audiences at the beginning of evaluations. When commissioning and impact evaluation it is important to register the design at the beginning at 3ie or other relevant registries. Then as evaluation outputs are generated these should be packaged and shared.

Roles and Responsibilities

6.12 Each programme should evaluate interventions in ways that maximise both the project/programme and higher-level learning opportunities. As evidence for Payment by Results is generated DfID will consider how best to facilitate cross-programme and cross-instrument synthesis and comparisons using central or external support functions.

6.13 Within DFID there are clear roles and responsibilities for staff depending on their function and position in the organisation:

- **DFID's Programme Teams:** The DFID team that designs and manages a Payment by Results contract is responsible for identifying evaluation needs and budgeting for associated costs. During implementation this team has primary responsibility for managing the evaluation of their programme;
- **DFID's (embedded) Evaluation Advisers:** Evaluation advisers embedded within DFID spending units are responsible for supporting the design and implementation of evaluations funded by programmes in their focus areas (geographic or thematic);
- **DFID's Evaluation Department:** The Evaluation Department sets evaluation policy and strategy and provides guidance and support to DFID's embedded Evaluation Advisers and other staff
- **DFID's Finance and Corporate Performance Department (FCPD):** FCPD supports the design and management of DFID programmes and assesses the value for money of payment by results aid modalities.

6.14 Each programme should clearly identify how their evaluation will be managed to uphold quality and independence. DFID recommends that evaluations are procured

separately from other payment by results activities, like outcome verification. This enables the independent assessment of unexpected outcomes associated with verification.

7. Conclusions and Next Steps

7.1 The purpose of this Evaluation Framework is to guide Payment by Results evaluations towards building evidence for what works best in different contexts and for different purposes. The guidance is focused on the Payment by Results instruments and not intended to replace guidance for programme evaluations.

7.2 DFID and many other donors are using Payment by Results because they believe that in some circumstances these instruments will have positive outcomes. Expected positive outcomes from using Payment by Results include: 1) Increased transparency and accountability for results and stronger focus on performance; 2) Increased potential for innovation and flexibility in delivery; 3) Improved efficiency and effectiveness of delivery; 4) Interventions achieve intended outcomes and impacts.

7.3 Given the early state of evidence, DFID encourages Payment by Results programmes to consider what evidence can be generated about both the performance (process) and impact of the instrument. Many evaluations are likely to focus on the performance and impact of an intervention for programme learning and accountability purposes. DFID encourages that, in addition to generating evidence for programme purposes, Payment by Results evaluations also contribute to learning about the instrument effects of Results Based Aid, Results Based Finance, and Development Impact Bonds.

7.4 Generating evidence for Payment by Results requires designing evaluations for learning about the intervention and also contributing evidence for comparisons and accumulation. Mixing methods can be a useful way of capturing both 'if' and 'how' selected types of Payment by Results have any intended or unintended outcomes. Evidence generated about Payment by Results should be actively communicated and synthesised.

Department for International Development

22 Whitehall
London
SW1A 2EG, UK

and at:

Abercrombie House
Eaglesham Road, East Kilbride
Glasgow
G75 8EA, UK

Tel: +44 (0)20 7023 0000

Fax: +44 (0)20 7023 0016

Website: www.dfid.gov.uk

Facebook: www.facebook.com/ukdfid

Twitter: @DFID_UK

Email: enquiry@dfid.gov.uk

Public enquiry point: 0845 3004100 or +44 1355 84 3132 (if you are calling from abroad)

© Crown copyright 2014

Copyright in the typographical arrangement and design rests with the Crown. This publication (excluding the logo) context. The material must be acknowledged as Crown copyright with the title and source of the publication specified. Cover image © Russell Watkins/Department for International Development
Published by the Department for International Development, May 2014