

Verification in Practice #3:

What makes a good indicator for a Payment by Results programme?

Lessons learned from verifying the DFID WASH Results Programme

May 2020

This practical note has been written to help those involved in commissioning or managing Payment by Results (PbR) programmes to set and agree good quality indicators. Produced by e-Pact, the Monitoring & Verification (MV) team for the DFID WASH Results Programme, the note suggests seven essential characteristics of a good PbR indicator and five additional characteristics that may be desirable.

What's different about PbR programme indicators?

Good PbR indicators are similar to any other good indicator, the main difference is that, being the mechanism for translating payment-linked results into practice, more rests on getting them right in PbR programmes. Progress against indicators will determine how much suppliers get paid; if progress cannot be accurately measured then contractual agreements may need to be revisited. In PbR programmes, indicators are one of the main levers that commissioners

have for shaping what programmes do and the quality of implementation; if commissioners choose the wrong indicator or measure it in the wrong way, they may not get their intended results or may pay more for achieving them. In this note, we present what we believe to be best practice in designing PbR indicators. While we use adapted examples from the WASH Results Programme to illustrate our discussion, this should not replace the detailed design work that is essential as part of any PbR project.

“Agreeing results to be targeted and choosing appropriate indicators is fundamental to PbR, as what gets measured (and rewarded) gets done.”

Designing and Delivering Payment by Results Programmes: A DFID Smart Guide.

Box 1: Results, indicators and targets in Payment by Results

Results detail the intended areas of achievement; in the WASH Results Programme these are outputs and outcomes in water, sanitation and hygiene, e.g. “Number of people using an improved sanitation facility”. In many other results or outcome-based programmes these are known as “outcomes”.

Targets are the specific values attached to each result, for the purpose of determining whether satisfactory performance has been achieved, e.g. “300,000 people with access to clean drinking water”. Targets are usually, but not always linked to payment. In other programmes, these are known as “outcome metrics”, “triggers” or “disbursement-linked indicators” (DLIs).

Indicators are the way of measuring progress towards the outcomes or results. e.g. “Survey respondents mention at least two ‘critical times’ for handwashing AND where the household has access to a handwashing device within 10m of the toilet where soap and water are present”. In some programmes they are known as “outcome measures”.

This *Verification in Practice* note focuses on indicators, not targets or metrics. For a more detailed explanation of the differences between them, see the Go Lab guide to setting and measuring outcomes, available at <https://golab.bsg.ox.ac.uk/guidance/technical-guides/setting-and-measuring-outcomes/>.

Why is it important to agree indicators early in a PbR programme?

In many programmes, indicators are agreed, revised and rejected as the programme goes on. In PbR, the pressure to get them right at the beginning is intense because, together with results and targets, they inform the value of the contract. The indicators and means of measurement will also shape the nature and size of the Monitoring and Verification effort. This pressure to get indicators right at the beginning of a programme is compounded when some payments are linked to achievement of outcomes, as it was in the WASH Results Programme. Outcome achievements (e.g. proportion of people still washing their hands) are often measured by comparison with a baseline survey, so indicators and measurement approaches need to be agreed early on to ensure comparability of results.

Essential characteristics of PbR indicators

1. Reduce ambiguity
2. Well-aligned
3. Specific and clearly defined
4. Used in combination
5. Feasible to measure
6. Efficient to measure
7. Gender sensitive

Desirable characteristics of PbR indicators

8. Supportive of downward accountability
9. Contextually relevant
10. Clearly mapped to national and global indicators
11. Disaggregated
12. Aid learning and course correction

It is **essential** that PbR indicators are:

1. Capable of unambiguously informing payment decisions. When payment is dependent on an indicator, there is no space for ambiguity. Coupled with the quantitative, target-based nature of PbR programmes, this can drive a preference for indicators that are easy and reliable to measure, even if they are not the best measure of (or proxy for) the programme's intended outcomes and impacts. This is challenging when the intended impacts of a programme rely on changes in behaviours, systems and relationships that can be difficult to measure quantitatively. The WASH Results Programme's emphasis on assessing outcomes required detailed consideration of how to measure sustained use of latrines, sustained access to water, and handwashing behaviour in ways that generated unambiguous figures. Indicators and their measurement had to be well-designed to assess qualitative changes in quantitative terms. One strategy for setting indicators for hard-to-quantify progress in areas such as "improved sector alignment" is to include secondary indicators that are not payment-linked, or indicators for which payment is linked to the reporting, not the progress (see Box 2 on sustainability indicators).

2. Well-aligned with intended results. Indicators need to be carefully selected to incentivise the desired results; setting indicators is one of the commissioners' key levers for determining the nature of programming. Poorly chosen indicators can inadvertently generate perverse incentives. For example, if one wanted to change hygiene behaviours, an indicator based on the presence of a handwashing station could encourage the provision of materials to construct cheap, inappropriate and unsustainable handwashing stations, without a focus on changing hygiene behaviours. Similarly, choosing an indicator of people reached with hygiene promotion, measured using hygiene promotion records could encourage the cheapest form of hygiene promotion to reach the most people, rather than the desired outcome of sustained hygiene behaviour change. By contrast, one progressive outcome indicator used in the WASH Results Programme was "population practising handwashing with soap at critical times" which prompted serious thinking about how to measure hygiene practice and made sure that the critical importance of timing of the practice was considered (washing hands before eating being more important than afterwards).

3. Specific and clearly defined. With PbR, the devil is in the detail. It is important to agree upfront exactly what is meant by terms such as “improved”, “access to” or “reached” rather than leaving these terms to be defined once the programme is operational. The definitions used could influence suppliers’ decisions about the nature of programming, the approach to measurement and ultimately how much suppliers will be paid. The definitions adopted can help to outline expectations on quality of programming. For example, an indicator around increasing the number of people with access to a latrine that does not specify a maximum number of users for each latrine risks incentivising unsustainable and ineffective programming. Agreeing that maximum number will require consultation with local communities (e.g. around average household size) and a contractual discussion between supplier and commissioner because the agreed figure may have financial implications for either or both.

4. Used in combination. Often, more than one indicator is required to take a nuanced approach to assessing progress towards an intended result. In one supplier’s programme, three indicators were agreed to assess handwashing behaviours based respectively on: knowledge, reported practice and observation of facilities. This provides a potentially strong way of understanding the extent of behaviour change. In this instance, each component indicator accounted for one third of the outcome payment independently of the others. This meant, for example, that it was possible to achieve (and be paid for) the target of *reported practice* of handwashing with soap, even if this was not matched by a corresponding increase in the presence of soap. When multiple indicators are used within a PbR programme, it is important to consider how data from each indicator will be aggregated to generate a “result”, ensuring that no perverse incentives are generated and, if necessary, adjusting the targets appropriately.

5. Feasible to measure. Good practice suggests that an indicator should only be selected if it is possible to reliably collect data to measure progress towards it over the whole period in question (in this case the duration of the WASH Results Programme). However, data collection to measure indicators in the WASH Results Programme has been impeded for diverse reasons such as elections, a nationwide census, an earthquake, flooding, seasonal factors and conflict. This is more important in a PbR programme because failure to collect data may impact on payment, either in the short term if results cannot be evidenced in time, or in the longer term if base, mid or endline surveys are affected, making changes in outcomes difficult to quantify. In fragile contexts, plans may need to be made for remote data collection and verification – it may be necessary to make trade-offs between rigour, practicality and cost of verification. Discussions about risks to data collection need to happen during the design phase.¹

6. Efficient to measure. In a PbR programme, reliable data needs to be collected for every payment-linked indicator agreed. To avoid putting undue burden on frontline staff, every effort should be made to identify indicators that can be measured using data already being collected by suppliers for programme monitoring purposes. The effort of gathering evidence of sufficient rigour for payment needs to be proportionate to the value of the indicator. For example, one output indicator on hygiene promotion required the collection of detailed evidence of attendance at community events, but that quickly became unwieldy and was subsequently revised. Such examples led some stakeholders in the WASH Results Programme to refer to the programme as “payment by paperwork” and the independent evaluation of WASH Results Programme to suggest that process-related indicators could be de-linked from PbR, since they were time consuming to document and verify.

7. Gender sensitive. For example, the handwashing with soap outcome indicator ***“Percentage of people who can name three or more critical times for handwashing increases by 15 percentage points at endline compared to the baseline”*** has significant gender complexity, as three of the five critical handwashing times concern childcare or food preparation. If men are typically not involved in these practices, then they will be less likely to correctly identify these as critical times. If the indicator does not recognise this fact, suppliers’ achievements may not be fully recognised. Introducing a gender-sensitive indicator (where men only have to identify two critical times) more closely matches the behaviour the WASH Results Programme is aiming (and able) to influence.

¹ <https://washresultsmve.wordpress.com/2018/05/16/truly-exceptional-handling-misfortune-within-payment-by-results/>

It is **desirable** that PbR indicators are:

8. Supportive of downward accountability. PbR has been accused of promoting upward accountability and, if this is a concern, we believe it could be counterbalanced by using indicators set by beneficiaries and/or including a requirement for beneficiary feedback as part of verification of results, e.g. through satisfaction surveys, scorecards or focus groups.²

9. Contextually relevant. Detail of indicators may need to reflect contextually specific behaviours such as water carrying or soap substitution or conduct of a critical practice in a different place to that envisaged by the survey designer, e.g. ***“The interviewers should observe if soap or a soap substitute is available at their handwashing facility.”*** However, this may make data harder to aggregate, particularly in multi-country programmes, and risks incentivising sub-optimal or risky programming if suppliers choose indicators that incentivise practices which may be less effective, such as use of ash rather than soap.

10. Clearly mapped to national and global indicators. Programmes may choose (as above) to design indicators that are specific to the contexts in which they work but should try to ensure that these can map clearly and unambiguously to accepted national indicators (which in turn map to global indicators). This can help ensure data is of use to local authorities, and may also make aggregation easier.

11. Disaggregated. Designing indicators that refer to specific groups, usually more vulnerable sections of communities, can be a way of incentivising equity and inclusion in programming and a way to mitigate risk of so-called “cherry picking” by suppliers, where they keep costs down by targeting the easiest to reach. However, there are implications for measurement as disaggregation may require an increase in sample size to provide results which are sufficiently statistically significant to use as a basis for payment decisions.

12. Useful for learning and course correction. Some WASH Results Programme Suppliers have argued that the emphasis on rigorous monitoring, reporting and verification of progress towards indicators has been at the expense of learning processes. Data collected to evidence indicators should also enable learning, thus an important consideration when selecting indicators is how meaningful and useful they are to the people implementing programmes. Data and information on progress and results need to be readily available, so that they can feed into adaptive management processes and enable course correction.

For the desirable indicators it may not be possible to achieve all of these simultaneously. For individual programmes there may be trade-offs between, for example, making indicators contextually relevant whilst attempting to map them to global indicators.

² <https://washresultsmve.wordpress.com/2016/11/10/can-payment-by-results-raise-the-bar-for-downward-accountability/>

Box 2: Emerging insights on sustainability indicators

The WASH Results Programme aims to achieve sustainable WASH outcomes that endure beyond the lifetime of the programme. Consequently, much thought has been given as to how sustainability can be incentivised within a PbR framework. This is challenging, first because it is difficult to measure (and pay for) future sustainability within the lifetime of a programme; second because many of the factors that ensure sustainability are not immediately within the control of suppliers so are risky to link to payment. It has been difficult to identify indicators which provide a useful measure of “how likely it is that use of facilities (or practice) will continue in future, and that systems and capacity are in place to support these sustained practices”. Any such indicators are, by definition, proxies and there is limited evidence in the sector as to what the most reliable proxies are. Many of the factors that underpin sustainability are difficult to measure quantitatively, rigorously and proportionately. For example, one supplier observed:

“Many of our sustainability indicators are based on joint reflection with partners about their capacity, using a scorecard. For example, with a group of stakeholders in a district, to discuss alignment, joint planning, inclusion. With PbR, we had to put much more rigor into these reflections, e.g. getting participation lists and signatures for each reflection in each district...The bigger challenge was that many stakeholders would rate their capacity high at baseline (depending on the country) and that in some cases political changes led to a complete change in staffing from baseline to next measurement.”

Basing payment on such indicators is challenging. DFID has responded flexibly to these challenges, in some cases linking payment to reporting on sustainability indicators rather than progress against them. Risk can also be managed through limiting the value of payments linked to sustainability indicators. There is emerging consensus that the requirement to report on sustainability has focused attention on what is meant by sustainability and how it can be promoted and monitored in WASH programming, so generating learning that is valuable for the sector.

Further reflections from the MV team on PbR indicators

- **Well-designed indicators can help to incentivise achievements and promote high quality programming.** However, there are ongoing challenges within the WASH Results Programme in striking the balance between setting meaningful indicators that reflect programme ambition and context, while also ensuring that monitoring them is not overwhelming and they can be aggregated at programme, donor, national and global levels.
- **Indicators that appear straightforward and possible to measure may not be in practice,** and some iteration may be required. A “dry run” of a verification round (that is not directly linked to payment) during an inception phase may help to identify indicators that are not working as planned.
- **Changes to indicators and their measurement could impact on the results achieved,** for example a shift to a less rigorous measure may generate higher reported results. The implications of changes to indicators and their measurement need to be carefully considered in a tripartite process between the commissioner of the PbR programme, supplier and MV team, before changes are agreed between supplier and commissioner.
- **Linking payment to indicators has served to focus attention both on the outcome being measured and the means by which it is measured,** so contributing to better thinking and practice in the sector. Our emerging experience suggests that linking even modest payments to indicators can have this effect. This can be useful for indicators linked to sustainability or other desirable activities – whether that be systems strengthening or ensuring the inclusion of marginalised groups in programming – where there is uncertainty about what outcomes it is reasonable to expect or where statistically representative measurement and verification is not possible.

Further information:

About the WASH Results Programme

(*Water, sanitation and hygiene results programme to support scale-up efforts, [GB-1-203572]*)

The WASH Results Programme aims to support poor people in 12 countries to access improved water and sanitation, and to introduce improved hygiene practices. Three NGO consortia ('Suppliers') were contracted by DFID to reach 4.9 million people, initially. In response to DFID's commitment to the Sustainable Development Goals and as part of DFID's strategy to tackle extreme poverty by delivering the 'basics' of development, including extending access to clean water and sanitation, WASH Results was expanded to ensure an additional 2.5 million people gain access, to be completed in 2021.

About the authors of this *Verification in Practice* note

This note was produced by the WASH Results Programme's Monitoring and Verification team which is composed of staff from the e-Pact consortium (Itad, IWEL, OPM, and Ecorys). It was produced originally for DFID in 2018 and subsequently edited for wider audiences. Drafting and editing was led by Cheryl Brown and Catherine Fisher, with guidance, comments, corrections and specific content from Alison Barrett, Don Brown, Joe Gomme, Ben Harris, Andy Robinson, Amy Weaving and Kathi Welle. We are particularly thankful to Stephen Lindley-Jones and Anne Joselin at DFID for their support and to the DFID reference group who gave useful feedback on an earlier draft.

More publications by the WASH Results Programme's MV team on verification of PbR:

- *Verification in Practice #1: The Verification Cycle: Step by Step*
- *Verification in Practice #2: Appraising Monitoring Systems*, e-Pact, 2020
- *DFID Payment by Results Guidance Note: Lessons from an effective verification system*, e-Pact, 2020
- Sarah Holzapfel Heiner Janus (2015), *Improving Education Outcomes by Linking Payments to Results: An Assessment of Disbursement-linked Indicators in five Results-based Approaches*, Discussion Paper / Deutsches Institut für Entwicklungspolitik

Related online articles and blogposts

Measuring progress towards SDGs: a Payment by Results perspective

<https://washresultsmve.wordpress.com/2017/08/09/measuring-progress-towards-sdgs-a-payment-by-results-perspective/>

How can we develop or select good performance indicators?

https://www.betterevaluation.org/en/resources/ANZSOG/develop_select_good_performance_indicators



e-Pact is a consortium led by Oxford Policy Management and co-managed with Itad