Module 6: Evaluate

Structure of the Module:

Up until this point we have described how to plan, design and implement a hygiene behaviour change intervention. This last module describes how to evaluate a hygiene behaviour change program. Even though this module comes at the end, evaluation is actually something that needs to be considered at every stage of the ABCDE process. In this module we outline the various different things that can be learned from an evaluation and how they can best be measured. This module is broken into the following sections:

Key learning points:

- To design an evaluation, it is important to understand the purpose of the evaluation. To do this we identify the stakeholders of a program might and consider the kinds of questions they might want to answer.
- We often assume that the only thing that needs to be assessed is whether our program has worked (either to change behavioural or to improve health). We call this kind of an evaluation an Impact Evaluation.
- Impact evaluations cannot, however, tell you why an intervention works. For this we need a Process Evaluation which explores aspects of the intervention's theory of change to understand whether the intervention was delivered as intended.
- WaterAid does not recommended that we measure health outcomes as these require complex and expensive study designs. Therefore, WaterAid programs will primarily assess behavioural outcomes.
- There are three main ways to collect data on behaviour – through observation, spot-checks and self-report. Self-reported data on behaviour is limited since people often do not tell you what they actually do. Observation is the gold standard as it truly captures behaviour, but is hard to implement at scale. WaterAid recommends using spot-checks as the primary method of data collection and where possible this should be supported with observation.

Supporting Resources

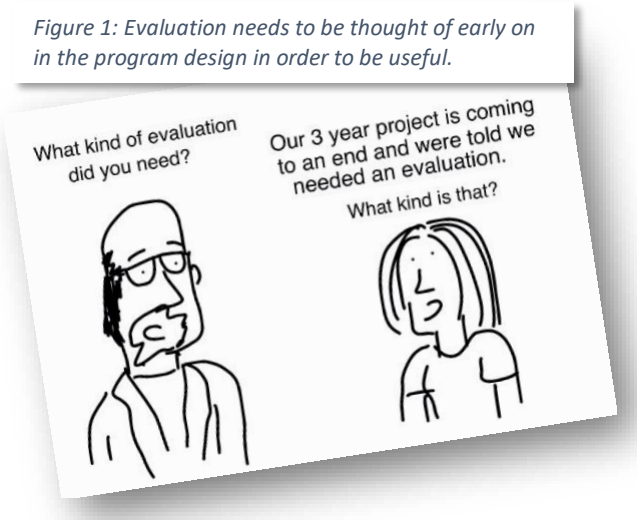
- Example of a structure observation form

Why do we need to evaluate our hygiene promotion programmes?

The purpose of evaluations is to influence future decision making. Generally, we want to evaluate programs in order to do one of three things:

- To decide whether to continue the existing programme or not.
- To decide whether to change or redesign the programme
- To decide whether to roll out a similar programme elsewhere.

Figure 1: Evaluation needs to be thought of early on in the program design in order to be useful.



Many different people and organisations may be invested in a WaterAid programme. These individuals may have different expectations about what they will learn from an evaluation, so it is important that you understand who the evaluation is for and how the findings will be used before you plan the evaluation. To decide on the purpose of an evaluation we need to think about who the target audience is for our evaluation and consider what they want to know, and what action they will take as a result. Table 1 lists some of the key questions that an evaluation may seek to answer and some of the different audiences who may be interested in the answers to these different questions.

Table 1: Factors to consider when deciding on the purpose of an evaluation

Audience	Questions of interest	Potential action to be taken
Programme managers, administrators, funders	Is the programme performing as expected?	Improvement of programme
Administrators, funders	Is the programme worth continuing?	Decision on continuation of funding
Administrators, funders and policy makers	Should it be extended?	Decision on replication of programme elsewhere
Researchers, policy makers	Is it causally linked to improved health outcomes (e.g. reduced diarrhoea)?	Link improvements in behaviour or health to the programme

Additionally, WaterAid programs are delivered through ongoing cycles which allow us to learn from our experience. Figure 2 shows how planning, monitoring and evaluation all contribute to the way WaterAid programs are implemented.

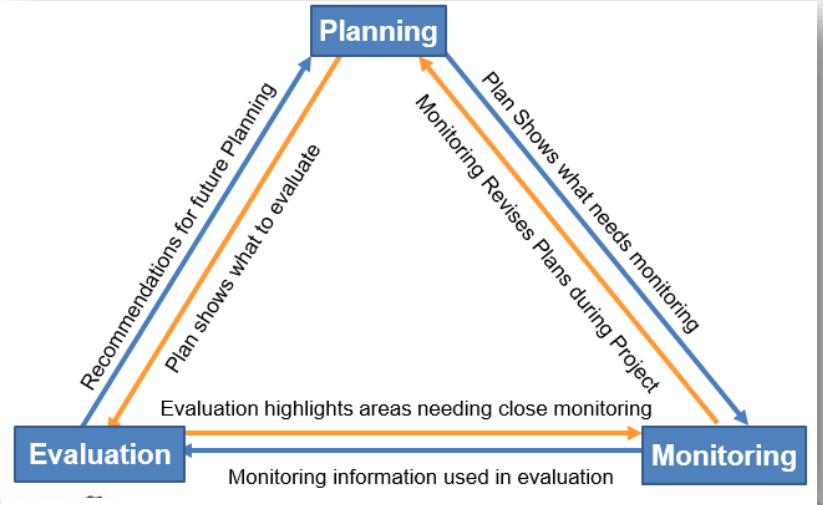


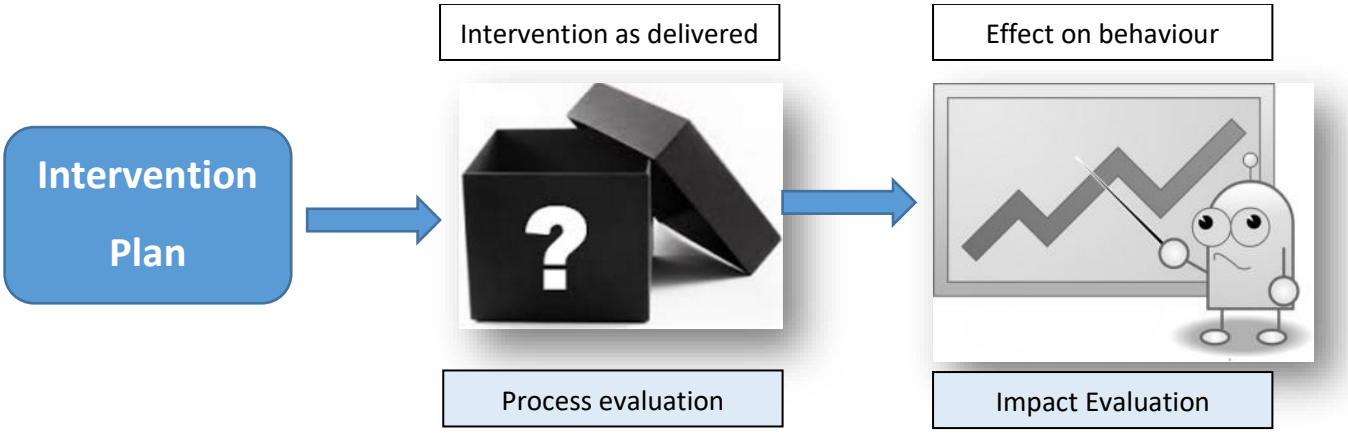
Figure 2: WaterAid designs it's programs using a cycle approach so that planning, monitoring and evaluation processes are interconnected.

What do we want to find out in our evaluation?

Once we know the purpose of our evaluation we can think about what it is that we want to assess. In Figure 2 you can see four evaluation areas and the links between the programme and the impact it hopes to achieve.

1. A programme needs to be delivered as intended **(provision)**
2. The target population needs to participate in the programme **(coverage)**
3. Participation in the programme needs to result in practice of the desired hygiene behaviour **(compliance, or behavioural impact)**
4. Practice of a desired hygiene behaviour should lead to a health impact **(health impact)**

Figure 3: An impact evaluation measures the effect of an intervention on behaviour. However if we don't do a process evaluation to fully understand what actually happened when the intervention was implemented we not understand why the behaviour change occurred.



Assessing provision and coverage can help us to understand how a programme has performed. This is often called **process evaluation.** Investigating whether or not a programme has worked (it's impact on behaviour or health) is known as **impact evaluation**. While an impact evaluation can tell us whether or not a programme has succeeded in its goal of improving hygiene behaviour, it doesn't help us to understand why the programme has succeeded or failed. This is where process evaluation comes in, as monitoring provision, utilisation and coverage can help you to answer questions about how to improve programme design and intervention delivery which may determine a programme's future. We will consider each of these evaluation areas in turn.

Measuring Provision

Investigating provision involves understanding whether the intervention implemented as planned? This refers to both the quality and the quantity of the programme delivered and is often called **fidelity**. Figure 4 provides an example of some of the different things that could be measured when assessing provision, although it is important to remember that this checklist would need to be tailored to your programme. Using a checklist or standardised template to assess provision of your programme is particularly useful for assessing quality. If you don't have a standardised approach you run the risk of one assessor deciding the programme was delivered

Figure 4: Key questions to ask for measuring provision

- ✓ Did session take place? How many times?
- ✓ Was the session carried out by the correct number of trained implementers?
- ✓ Were all necessary materials available for use in the session?
- ✓ Were materials distributed to participants (if relevant)?
- ✓ Did all elements of the session take place?
- ✓ Did implementers answer questions competently?
- ✓ Were there any logistical challenges that affected the session (weather, time delays technology, etc.)?

with “high” quality, while another evaluator may think there is considerable room for improvement.

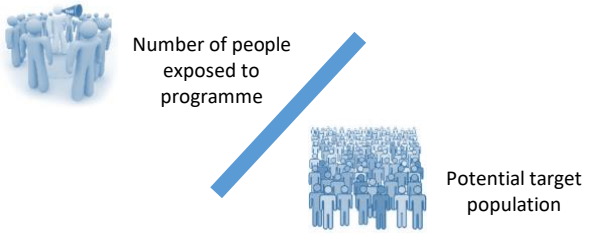
Measuring Coverage

Even if a programme has been delivered as intended you still need the target population to participate in the programme if you wish to change their behaviour. Assessing coverage assesses the extent to which the target population reached by the programme. To answer this, you need to assess **exposure** to the programme and you need to know or estimate how many people there are in the target population.

Exposure: You can ask people questions about the hygiene programme to find out whether they have attended it. You want these questions to be specific enough to your programme so they do not mistakenly tell you about another hygiene programme in the community. You might do this by asking questions about attendance at specific events run by specific individuals. If the programme had a brand or character associated with it, you can even show a picture or campaign logo when you ask about attendance at the event.

Target population: It can often be difficult to estimate the total population in our programme area. This often means that we are left the choice of using the data from an outdated government census or doing our own estimate which will take time and money.

Once you have an estimate of the number exposed and the total target population you can estimate coverage using a simple formula: the number exposed is divided by the target population.



EXAMPLE: Calculation of Coverage
1200 individuals were served by programme in a community with 3000 individuals.
The coverage of the programme in the community is 1200 / 3000 = 0.4 (40%)

A related area that you may also wish to measure is the **intensity of contact** with the target population. If a hygiene programme is designed to be delivered over 10 weeks, with one event per week, and individuals in your target population have only attended one or two of these sessions, then you would want to know this as it could affect whether or not your programme succeeds in improving hygiene behaviours. To understand this you would want to calculate how many of the people exposed to the programme had the intended number of contacts with the programme.

Measuring an Impact on behaviour

If a programme is delivered well, and high coverage of the target population is achieved, you now want to measure whether or not people are practicing the desired hygiene behaviour promoted by your programme.

There are many different ways of measuring behaviour, each of which has its advantages and disadvantages. If we understand the differences between the measures we are in a better position to select an appropriate way of measuring behaviour in our programmes. The three main ways of measuring behaviour are observation, spot checks and self-report.

Observation

- Example indicator: % of caregivers observed washing hands with soap at critical times (e.g. before food preparation)

Structured observation is often referred to as the "gold standard" for measuring behavioural outcomes. *Observation* involves watching what people are doing, and the *structured* part means that you don't watch everything, but that you are interested in recording only select activities. It is usually carried out in the early morning as that is usually when hygiene behaviours that we are interested in observing are carried out. Observation is considered the gold standard because it allows you to document actual behaviour as it occurs in a natural setting. Additionally, it can tell you about the behaviour of each individual family member, which is useful if you want to know, for example, if all individuals wash hands or only some. Importantly observation is an objective measure of behaviour. It is a fact and not a decision - either you see someone cleaning the toilet or you don't so what is recorded by enumerators should always be the same.

Figure 5: example of a person collecting data through structured observation in a household in India.

One downside of observation is that it is susceptible to **reactivity**, meaning that people can change their behaviour because someone is watching them. Reactivity can be reduced by:

- Not telling participants exactly what behaviour you are observing (instead you can say you want to learn about their daily routines).
- Choosing enumerators who are young local women so that people feel comfortable in their presence.

Figure 6: Example of what you might see during a spot check if you were assessing whether there was soap and water at the latrine.

Observation is time-consuming and can be costly because people need to spend several hours in each household and they require training. It is unlikely that you will usually use observation to evaluate behavioural impact across an entire programme as it is not feasible to carry it out on a large-scale, but it is certainly possible to do observation among a sub-sample of your target population.

Spot checks

- Example indicator: % households with soap and water present at the designated place for handwashing

A spot check also involves observation, but it is literally just a check, rather than observation over a period of time. Spot checks are also objective outcome measures, but you need to think about what the data you collect through spot checks tells you. For example, if a household does not have soap and water at the handwashing place then there is a good chance that household members are not

washing their hands with soap, but if soap and water are there then you cannot conclude that they are actually using it to wash hands at the critical times. This is why spot checks are said to be a proxy measure of behaviour – they indicate that behaviour may take place but they are not measuring actual behaviour or documenting who is doing it. Spot checks are ought to be the dominant method used by WaterAid to measure behavioural impact throughout their programs.

Self-report

- Example indicator: % target population able to recall 5 critical times for handwashing with soap

Figure 7: Example of an enumerator asking about someone's behavior as part of a survey.

Self-reported questions in surveys or interviews are the easiest way of capturing large amounts of data quickly and relatively cheaply, which makes them popular. However, one of the main issues with collecting reported data is that it is a subjective measure, unlike observation and spot check. What someone says they do is not a measurable fact as people say what they think you want to hear or what they wished they did which can result in an over-estimation of the practice of "good" behaviours. This is known as ***social desirability bias.*** Self-reported questions can be useful for measuring knowledge (like the example indicator) and attitudes or norms. WaterAid programs should try to avoid using self-report as a primary measure of behaviour.

Table 2: Summary of the main approaches used to measure behaviour

	 Observation	 Spot check	 Self-report
Pros	Objective measure Captures actual behaviour of individuals	Objective measure Quick and easy to add to surveys	Easy way of capturing large amounts of data quickly and at low cost
Cons	People can change behaviour Takes time Need trained staff (costly)	Is a proxy (doesn't measure actual behaviour or who is doing it)	Prone to bias (reported hygiene behaviours often very different to actual practices)
When to use	On a small-scale (sub-sample?)	In surveys of all sizes	To learn about knowledge or attitudes, or to document exposure to a programme
When to avoid	Large-scale evaluation	To measure actual hygiene behaviour	To measure actual hygiene behaviour

Measuring Acceptability

If a programme isn't acceptable to the programme staff then it is likely that they will not deliver it well. If the programme is not acceptable to the target population then it is likely that they will not attend the programme, or if they do attend the programme, they will not adopt the hygiene behaviour promoted by your programme. It is often helpful to assess acceptability qualitatively through interviews or focus group discussions.

Measuring an impact on health

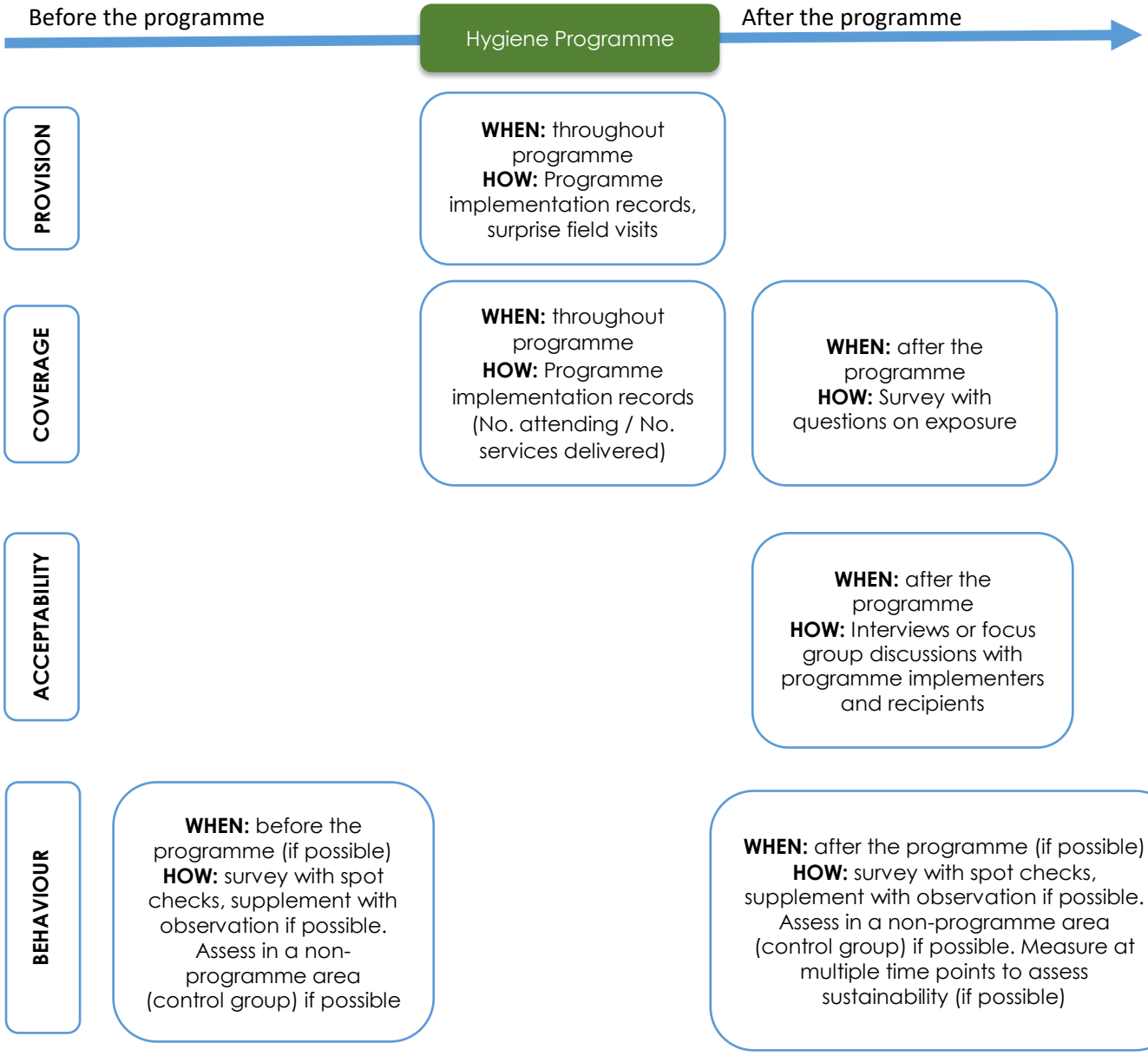
Even though our programme may be designed to improve health in some way, it is not usually possible to measure the impact of our programme on health. This is because it requires a complicated study design and a lot of resources.

When do we collect data for our evaluation?

- Data on **provision** of your programme should ideally be collected throughout programme implementation. You may not need to collect too much information specifically for the evaluation, as you should be able to make use of programme implementation records, but it can be helpful to conduct surprise field visits to observe the programme in action.
- Data on **coverage** can also be collected throughout programme implementation if your programme documents the number of people attending an event or receiving your programme. You can also conduct a survey at the end of your programme to measure exposure and calculate coverage from this survey.
- **Acceptability** can be assessed at the end of programme implementation in interviews or focus group discussions. A programme should always be piloted on a small scale so you can avoid implementing it if it is unacceptable.
- The simplest, cheapest way to collect data on **behavioural outcomes** is to conduct a survey at the end of programme implementation. If a survey is also conducted before the programmes starts (i.e. the evaluation is planned at the same time as the programme is designed) then you can measure behaviour before and after your programme which is a useful way of assessing whether behaviour has changed. To be more certain that change in behaviour following a programme is due to your programme you can also measure your behavioural indicators in a similar area that did not receive the programme.
- If you want to see whether or not improvements in hygiene behaviour are **sustained** over time you can go back and measure your behavioural outcomes months or years after the programme ends

Figure 8 gives a summary of the different types of data that you might want to collect and the timing of data collection.

Figure 8: Summary of evaluation areas and timing and nature of data collection



WaterAid's recommendations and minimum standards for Evaluation

Moving forward WaterAid want's its monitoring and evaluation approach to:

- Allow it to feel more confident about the impact that WaterAid is having in the world
- Design M&E programs with a clear purpose and target audience in mind so
- Use program evaluations to influence policy and the sector to the greatest extent by sharing what we have learned.
- Capture data on sustainability.
- Be able to draw cross-country comparisons between hygiene programs.

In order to achieve this WaterAid are recommending that as a minimum hygiene behaviour change programs use a standard minimum set of indicators. These have been chosen as a minimum as they

can be measured relatively rapidly and do not require extensive training of staff since they can all be assessed through a household spot check.

Table 3: WaterAid's minimum standard indicators for measuring hygiene behaviour.

Behaviour of interest	Indicator
Sanitation/excreta disposal	% of HH with human faeces visible present in the compound % of HH with a toilet that is visibly in use
Water Treatment/storage	% household storing drinking water in safe and clean container with lid % of HH able to demonstrate correct use of household water treatment equipment
Handwashing	% of HH with soap and water present at the handwashing facility
Menstrual Hygiene Management	% HH with menstrual hygiene materials available
Food Hygiene	% of HH with clean serving utensils at time of visit

Additionally, WaterAid recommends the following as a standard guide for all its hygiene programs.

Provision	Always measure so that you can tell why your intervention worked / did not work.
Coverage	Always measure so that you can tell why your intervention worked / did not work.
Data sources	Make use of available data wherever possible and if possible collect different types of data.
Data collection	Ideally, incorporate baseline and post-intervention data points to explicitly show improvements in target indicators (which means the evaluation needs to be planned before implementation starts
Health impact	Unless specifically requested do not measure health impact as a complex design is needed to draw any firm conclusions.
Observation	In addition to spot checks try to use observation among a sub-group of people in your program site.
Evaluation staff	Your evaluation team should not be involved in the intervention as this will bias the results.
Sustainability	Where budget allows, measure six months to one year after the termination of a programme to assess sustainability