

This picture reflects the past and the future of rural water supply in Zimbabwe. As the sector modernizes there is a shift from handpumps to solar-powered piped supplies. That is not only a shift in technology, but it also has implications for the service delivery model, on the financial arrangements and the role of local government.

But how can we measure systemic progress when change is complex, multi-layered, and unfolds over time?



IRC



© S. Smits 2024

Measuring What Matters

Rethinking Evaluation in WASH Systems Strengthening Programmes

By Stef Smits and Ruchika Shiva (IRC 2025)¹

Summary

Evaluating WASH (Water, Sanitation and Hygiene) programmes traditionally involves measuring tangible outputs - such as infrastructure delivered - alongside service-level outcomes and long-term health impacts. However, system strengthening programmes pursue a different kind of change: they aim to shift institutional roles, behaviours, and coordination mechanisms across the WASH system. As a result, they require a different evaluative lens - one that can capture complex, systemic transformations over time.

This paper presents the insights from the evaluation of Welthungerhilfe's Global WASH Programme (GWP2), which applied a systems strengthening approach across four countries.

The GWP evaluation combined process and effect evaluation using longitudinal, largely qualitative methods such as outcome harvesting and building block assessments. It reinterpreted OECD-DAC criteria - especially relevance, coherence, and impact - through a systems lens, focusing on contribution rather than attribution. A clear results chain was essential to track intermediate outcomes. The brief recommends that WASH systems programmes be evaluated based on their influence on structural, behavioural, and normative change - supported by embedded learning, adaptive management, and multi-stakeholder validation.

¹ACKNOWLEDGEMENTS: IRC deeply appreciates the contributions and active participation of the Welthungerhilfe country teams and local implementing partners who took part in the consultations, including: Water and Sanitation Promotion (WaSAP), Movement Towards Peace and Development Agency – Sierra Leone (MoPADA-SL), Farm Community Trust Zimbabwe (FCTZ), Mvuramanzi Trust (MMT), Association for Gender Awareness & Human Empowerment (AGAHE), Parmarth Samaj Sevi Sansthan, and Samarthan – Centre for Development Support.

²The Global WASH Programme, funded by the German Federal Ministry for Economic Cooperation and Development (BMZ), is implemented by Welthungerhilfe (WHH) across four countries: India, Pakistan, Sierra Leone, and Zimbabwe. The programme seeks to strengthen the long-term foundations for sustainable WASH service delivery by operating at three levels: the micro level (households and service providers), the meso level (districts), and the macro (national and state) level.

1 Introduction

1.1 Background

WASH programmes are increasingly shifting away from a focus on constructing basic water and sanitation infrastructure and implementing micro-level capacity-building measures - such as hygiene promotion or training of community-based management structures in specific locations - towards strengthening the WASH system as a whole. This shift reflects a growing recognition of the need to reinforce the systemic foundations for sustainable service delivery.

This evolution has significant implications for how progress is measured and evaluated. Traditional WASH evaluations have typically focused on tangible outputs (e.g., infrastructure delivered), immediate outcomes (e.g., improved access to services), and ultimate impacts on health and well-being. In contrast, WASH systems strengthening is not defined by a standardised set of interventions, nor does it lend itself to uniform evaluation metrics. Instead, it aims to influence the behaviours, roles, and interactions of multiple actors within the system - changes that

are often structural, incremental, and non-linear, and thus require different evaluative approaches.

System-level change in WASH service delivery is rarely attributable to a single intervention. It results from the interplay of multiple actors and contextual factors. As such, evaluations must focus on understanding the contribution of a given programme within a broader ecosystem of change, rather than attempting to isolate its impact.

In the absence of a standard methodology for evaluating WASH systems strengthening, Welthungerhilfe's Global Programme on WASH Systems Strengthening incorporated an evaluative research component. This component was designed not only to assess the outcomes and impacts of the GWP itself but also to generate methodological insights into how evaluations of systems-focused WASH programmes can be effectively designed and implemented.

1.2 Purpose, structure and audience of the paper

This paper aims to share the learnings on how WASH systems strengthening programmes can be effectively evaluated, by drawing on key lessons from the approach used in the evaluation of Welthungerhilfe's Global WASH Programme.

Section 2 outlines current approaches and methods for evaluating WASH systems strengthening, with particular reference to the literature review by Smits and Butterworth (2023). This is followed by a description of the evaluation design developed for the GWP and its application during the pro-

gramme's mid-term (Smits and Shiva, 2024) and final evaluations (Smits and Shiva, 2025). The final section synthesises key methodological lessons and reflects on their relevance for future evaluations of WASH systems strengthening initiatives.

While the paper primarily targets practitioners involved in managing, monitoring, or evaluating WASH systems programmes, its insights are equally relevant for those working on systems approaches in related sectors such as health, nutrition, food security, and environmental governance.

2 Evaluating WASH Systems Strengthening: Frameworks, Methods and Criteria

There is a growing body of literature on approaches, and methodologies used to assess WASH systems strengthening efforts. This section draws on the literature review by Smits and Butterworth (2023) to

present key conceptual foundations, including how WASH systems and their strengthening are defined, and how both the processes and outcomes of systems interventions can be evaluated.

2.1 Defining WASH systems and the logic of system strengthening

The WASH system can be referred to as “the actors and factors” (Huston and Moriarty, 2018) or “the people, behaviours, policies, processes, resources, interactions and institutions” (Casey and Crichton-Smith, 2020) necessary for delivery of WASH services. Ensuring that the WASH system can provide

sustainable WASH services requires a **systems approach**, or “...a philosophy of action, a way of working that recognises the complexity and fundamentally inter-linked nature of the real world” (Huston and Moriarty, 2019).

A systems approach is operationalised through **WASH systems strengthening**, which has been defined as “a process of analysis, implementation, adaptation and learning used to address the barriers to achievement of inclusive, sustainable, universal access to WASH - recognising that this access is the result of interactions between multiple actors and factors in a complex, dynamic system” (Casey and Crichton-Smith, 2020). Systems strengthening is implemented through a series of targeted **system interventions** - deliberate actions undertaken to strengthen specific components of the system. In most cases, a combination or **package of interventions** across different parts of the system is required to achieve meaningful and lasting change.

2.2 Evaluation Scope: What should be measured in systems strengthening?

When evaluating WASH systems strengthening initiatives, a critical first step is to clearly define what is being evaluated. Unlike conventional WASH programmes that focus on discrete outputs (e.g. infrastructure built, people reached), systems approaches operate through complex, multi-actor processes aimed at enabling long-term structural change. As such, the scope of evaluation must be adapted accordingly. There are two primary and distinct areas of evaluative focus that are recommended by this study:

1. The System Strengthening Process

This dimension assesses whether the programme has followed a coherent and effective process to strengthen the system. This includes evaluating how the programme identified system-level bottlenecks, whether relevant stakeholders were engaged, whether intervention logic was evidence-informed and adapted over time, and whether coordination with other actors was pursued. In essence, the process evaluation asks: Did the programme navigate the system effectively and strategically?

2. The Specific Systems Interventions

This dimension focuses on evaluating the relevance and effectiveness of the actual interventions implemented - whether they addressed the identified weaknesses in the WASH system and whether they contributed to observable changes in the system's functioning. This includes assessing both the design of the interventions (Were they the right ones?) and their results (Did they lead to systemic change?). This can involve changes in institutional behaviour, resource allocation, regulatory capacity, or service delivery norms.

In line with these two dimensions, four distinct types of evaluation may be applied in systems strengthening contexts, as recommended by De Savigny and Adam (2009), depending on the specific focus of the assessment:

- **Process evaluation** assesses whether the intervention was implemented as planned, and whether adjustments were made that influenced its effectiveness.
- **Context evaluation** examines changes in the external environment—political, economic, or institutional—that may have influenced the intervention's success or failure.
- **Effect evaluation** focuses on whether the intervention achieved its expected outcomes, such as improved system functioning or stakeholder behaviours.
- **Economic evaluation** investigates whether the intervention delivered value for money, weighing its costs against measurable or expected benefits.

Understanding the logic of WASH Systems Strengthening: The Role of results chains

To clarify the expected relationship between system interventions and their effects on the WASH system, the use of a results chain is common in systems strengthening approaches (Huston and Moriarty, 2018). A typical results chain starts with one or more system interventions (inputs), which contribute to a strengthened WASH system (intermediate outcome). This, in turn, leads to improvements in WASH service delivery for everyone (final outcome), as measured by relevant performance or service delivery indicators. These final outcomes are expected to contribute to broader societal impacts, such as improved health, enhanced livelihoods, and increased school attendance.

Box 1: The role of result chains

What should not be evaluated

Some calls have been made to evaluate whether a systems approach as such is more effective than conventional WASH programming. However, this is methodologically unfeasible. A systems approach is not a standardised model or fixed set of activities; it is a way of thinking and acting in complex, adaptive environments. As such, it cannot be evaluated by simply comparing it to a single alternative or control scenario - for example, what might have happened if a conventional project had been implemented instead (as in a counterfactual or control group design). Moreover, defining what constitutes a conventional approach is itself ambiguous - does it mean hardware-focused delivery, behaviour change campaigns, or community management models? The diversity of “non-systems” approaches makes such comparisons analytically weak and practically uninformative. Therefore, evaluations of WASH systems strengthening should avoid asking whether a systems approach works in general, and instead focus on how well a specific programme has operationalised systems thinking - through its process and interventions - and whether this has contributed meaningfully to strengthening the WASH system in context.

Box 2: Conventional WASH programming

2.3 Evaluating specific system interventions: Design, complexity, and methodological lessons

Many of the reviewed documents describe the assessment or evaluation of specific systems interventions, leading to the following insights:

- **System interventions are typically bundled into evolving packages.** Because WASH systems involve multiple interrelated factors, strengthening efforts usually require a combination of interventions targeting different parts of the system. This aligns with findings by Valcourt et al. (2020), who observed that many studies address several system factors simultaneously.
- **Interventions are rarely predefined in full at the outset.** Barely any cases predefined all the systems interventions at the start of a project. Instead, interventions emerged gradually based on ongoing analysis of context-specific gaps and needs. This adaptive approach is particularly common among organisations with long-standing engagement in a given region, where interventions extend beyond the scope of a single project cycle.
- **Evaluations tend to apply a longitudinal rather than a counterfactual design.** Nearly all studies examined system change over time within the same geographic area - comparing the situation before and after the interventions rather than comparing outcomes with a similar area where no intervention was implemented (as in a counterfactual or control group design). These longitudinal approaches typically capture: (1) a baseline assessment of the system, (2) a description of the interventions implemented, (3) experiences during implementation, and (4) observable results. Only one study (Stanford Program on Water, Health & Development, 2020) employed a counterfactual design using comparison districts.
- **Each step in the results chain must be monitored separately.** While most studies assessed outputs and contributions to intermediate outcomes, few systematically examined final service-level outcomes. This reflects the reality that no single method can capture the entire results chain - from inputs to long-term impacts. Instead, different tools must be used for different stages (e.g. output verification, outcome harvesting, service level measurement), acknowledging that causal links between steps are often non-linear and indirect.

2.4 Diagnosing system performance using building block frameworks

A core element of WASH systems strengthening is the ability to understand and assess how well the system is functioning. To strengthen a system effectively, one must first establish a clear picture of its current state - its capacities, weaknesses, and performance dynamics.

To this end, a number of organisations and authors have conceptualised the WASH system in terms of **building blocks** - discrete but interrelated components that together determine the system's overall functionality. These frameworks allow practitioners to assess the enabling environment for service delivery and identify which areas require targeted interventions. The concept of building blocks has been shaped by several major global initiatives, notably the World Bank's Country Status Overview series (AMCOW, 2011), the WHO-led GLAAS (Global Analysis and Assessment of Sanitation and Drinking-Water), and UNICEF's WASH Bottleneck Analysis Tool (WASH BAT) (UNICEF, 2018).

Building on these foundations, major WASH sector NGOs such as IRC, WaterAid, Water For People and Welthungerhilfe have adapted and applied building block frameworks in their programming and analysis.



Fig.1: WASH Systems conceptual framework with its 9 building blocks applied by Welthungerhilfe

A review of these frameworks (Battle, 2019) highlights several key features and considerations:

- **Semi-quantitative assessment methods** are commonly used. Each building block is assigned a score based on a set of criteria, with a qualitative justification provided to explain the rating. This mixed-method approach captures both measurable performance and contextual nuance.
- **National-level focus** remains dominant, given that many system functions - such as policy formulation, regulation, and financing - are defined at that level. However, some organisations are increasingly adapting these frameworks to sub-national contexts, such as districts or municipalities.
- **Diagnostic and monitoring tool.** While building block frameworks can, in theory, be used for ongoing monitoring, they are most often applied as one-off diagnostics. These assessments typically serve as a starting point to identify system gaps and guide the design of strengthening interventions.
- **Risks and limitations.** Two key risks should be noted. First, the way “ideal states” are defined for each building block can inadvertently promote rigid or unrealistic standards (e.g. requiring an “independent regulator” rather than a “clearly defined regulatory function”). Second, focusing on individual building blocks in isolation may obscure the interdependencies and dynamic relationships between components - undermining a holistic understanding of the system.

In conclusion, building block frameworks offer a valuable lens for diagnosing system strength and designing targeted reforms. However, their application must remain flexible, context-sensitive, and system-aware - avoiding overly prescriptive or siloed interpretations.

2.5 Applying the OECD-DAC criteria in systems contexts

The six OECD-DAC evaluation criteria - relevance, coherence, effectiveness, efficiency, impact, and sustainability - are widely recognised as the global standard for evaluating development interventions. However, these criteria were originally designed for project-based, output-focused initiatives. When applied to complex and adaptive programmes such as WASH systems strengthening, which operate through non-linear change processes and involve multiple actors, the OECD-DAC criteria - often rooted in linear logic models and clear attribution - require careful reinterpretation.

Welthungerhilfe’s Global Programme on WASH Systems Strengthening therefore applied a reframed interpretation of the OECD-DAC criteria to better capture the nature of systemic change (Table 1).

OECD evaluation criterion	Evaluation questions for WASH systems interventions and implications
Relevance	In view of the current state of the system, are the right interventions being undertaken? Requires getting insight into how interventions were identified and defined, on the basis of an understanding of the limitations of the WASH system, and the needs and priorities of key actors.
Coherence	How well do the interventions fit with efforts of others to strengthen the system? Requires understanding the extent to which the intervention aligns with the system strengthening priorities of the sector, as well as interventions undertaken by others.
Effectiveness	Are the interventions resulting in the expected change? Requires measuring outputs and outcomes as a result of the intervention.
Efficiency	Are the resources used in a prudent manner? Requires assessing inputs in relation to the results along the results chain, thereby acknowledging (OECD, 2019) that analysing efficiency from inputs to impacts along the entire results chain is methodologically challenging. Benchmarking is impossible, as systems interventions are often unique.
Impact	What changes is the intervention leading to in the long run? Requires understanding the results that are longer term or broader in scope. This revised criterion includes “...the holistic and enduring changes in systems or norms, and potential effects on people’s well-being, human rights, gender equality, and the environment...” (OECD, 2018). This impact definition aligns well with GWP’s impact definition (see Annex 2).
Sustainability	Are the changes likely to be sustained after the intervention and other external support? Requires assessing whether system-wide requirements are in place to sustain the changes to the system that have been achieved.

Table 1: Evaluation questions and implications for WASH systems interventions

For example, the criterion of **impact** is no longer limited to improvements in individual well-being (e.g. health or income), but is now interpreted to include structural or normative shifts - such as institutional reforms, changes in accountability practices, or the adoption of new service delivery models. This interpretation aligns with the revised OECD definition (2019) and with Welthungerhilfe's internal impact framework (see annex 2).

Similarly, the criterion of **coherence** gains particular importance in a systems context. It calls for evaluating how well an intervention aligns with - and adds value to - efforts by other stakeholders to strengthen the system. This is critical, as system-level outcomes can rarely be achieved in isolation.

The main differences between the traditional OECD-DAC approach and the requirements for evaluating systems strengthening interventions are summarised in the table below.

Dimension	Traditional OECD-DAC Interpretation	Systems Strengthening Approach
Level of analysis	Focus on project outputs and service-level outcomes	Focus on intermediate systemic outcomes and institutional change
Attribution	Emphasis on direct causal links between intervention and result	Emphasis on contribution within a broader ecosystem of actors and factors
Evaluation unit	Single project or intervention	Dynamic system context with evolving packages of interventions
Use of criteria	Standardised application cross projects	Reframed criteria aligned with system dynamics and complexity
Measurement	Linear results chain with predefined indicators	Non-linear pathways with multiple feedback loops and time lags

Table 2: Comparing OECD-DAC and systems strengthening evaluation approaches

3 Designing and applying a systems-based evaluation approach

This section outlines how the evaluation approach for Welthungerhilfe's Global WASH Programme was designed and applied in practice, drawing on the methodological considerations introduced in Section 2.

3.1 Welthungerhilfe's outcome and impact Framework

When evaluating systems strengthening interventions, it is essential to align the evaluation approach with the implementing organisation's understanding of outcomes and impact. Welthungerhilfe's revised Impact Framework—adapted to capture results specific to systems strengthening—is presented in Annex 2.

3.2 Considerations for the design of the evaluation approach

The aim of the GWP-evaluation was to get insights into the extent to which its system strengthening approach had contributed to achievement of its intended outcomes, and to understand the effectiveness and underlying drivers of the interventions applied.

To meet these objectives, a combined **process and effect evaluation** approach was adopted, allowing for a structured application of the OECD-DAC criteria through a systems lens:

- The **process evaluation** focused on the internal logic and operationalisation of the GWP, with particular attention to i) the relevance and coherence of the intervention logic in light of local system needs, ii) the extent to which the programme adapted its approach to align with government strategies and actions of other stakeholders, and iii) the quality and effectiveness of the process steps undertaken to identify and implement system-level interventions.
- The **effect evaluation** focused on i) the effectiveness of specific packages of interventions in contributing to targeted outputs and intermediate outcomes, ii) the extent to which the programme contributed to systemic impact, and iii) whether system-wide conditions are now in place to sustain the changes achieved.

To support both dimensions of analysis, a **longitudinal evaluation design** was deemed more appropriate than one based on geographic counterfactuals. Rather than comparing programme and non-programme areas, the evaluation assessed change over time within the same locations. This allowed for a more realistic exploration of contribution across the different steps of the results chain - from outputs, to intermediate system outcomes, and ultimately to long-term impact.

3.3 Evaluation questions

The combined process and effect evaluation was translated into a set of overarching and specific evaluation questions, building on the generic systems-oriented questions outlined in Table 1. These were slightly adapted for the midterm review (MTR) and the endline evaluation. A full list of the detailed evaluation questions is provided in annex 1.

3.4 Phasing of the evaluation

The evaluation of the GWP was carried out in three distinct phases, each building on the previous one and aligned with the evolving maturity of the programme:

- **Development of results chains:**

At the outset of the evaluation, the programme's results chains were not yet fully articulated. A key preparatory step was therefore to support the development of these chains to clearly define the expected outputs, immediate outcomes, and intermediate outcomes. This was essential for establishing a shared understanding of the programme logic and for framing the subsequent evaluation phases.

- **Midterm Review (MTR):**

The midterm review focused primarily on **process evaluation**, with an emphasis on assessing the relevance and coherence of the intervention logic, the appropriateness of the chosen system interventions, and the degree of alignment with sector priorities and stakeholder efforts.

- **Endline Evaluation:**

The final phase focused more strongly on **effectiveness and impact**—examining the results achieved, the contribution of the programme to system-level change, and the extent to which conditions for sustaining those changes are now in place.

3.5 Data collection and analysis

Both the midterm review (MTR) and endline evaluation (ETR) of the Global WASH Programme were conducted using a predominantly qualitative, contribution-focused approach. This involved a combination of desk reviews, remote and in-person interviews during country visits, and joint analysis with the GWP team.

Document Review:

The evaluation drew extensively on project documentation and monitoring outputs, including:

- **Programme documents**, such as feasibility studies, project proposals, results chains, and annual monitoring reports, which provided the basis for understanding the programme logic and implementation trajectory.

- **Monitoring tools and internal assessments**

The evaluation drew on a variety of monitoring tools and internal assessments to understand progress and system-level change. These included the annual application of the building block assessment tool in most countries, which helped track the evolving strength of key WASH system components (see also Section 2.4). Capacity assessments were conducted—primarily during the inception phase—to analyse the roles, capabilities, and institutional arrangements of district-level stakeholders through both qualitative and quantitative methods. Outcome harvesting was used at least twice in each country to systematically capture changes in stakeholder behaviours and institutional practices, and to assess the degree to which these changes could be credibly linked to GWP interventions. In addition, the evaluation considered project outputs such as district WASH master plans, training documentation, technical studies, and learning notes that provided evidence of operational progress. Finally, sector studies were reviewed to contextualise findings within the broader WASH system and policy environment of each country.

Stakeholder Perspectives:

To complement the document review, the evaluation included extensive engagement with stakeholders through key informant interviews, focus group discussions, and stakeholder meetings - conducted both remotely and in person during country visits. These consultations involved

- GWP programme staff and implementing partners, including project leads, MEAL officers, and technical specialists, as well as representatives from national and subnational government institutions such as ministries and local authorities. Additionally, the evaluation team engaged with service providers, private sector actors, and civil society organisations involved in WASH service delivery at the district level. Community perspectives were also captured through discussions with local leaders and beneficiary groups in selected villages, providing insights into micro-level changes and reinforcing feedback loops between system actors and end users.

Joint Analysis and Validation:

To ensure credibility and shared ownership of findings, validation workshops were conducted with GWP country teams at the end of each evaluation phase. These sessions provided an opportunity to review preliminary findings, test assumptions, clarify contextual factors, and reflect collectively on the implications of the evaluation. Additional validation meetings were held with the GWP global coordination team, including the WHH WASH advisor, programme coordinator, and MEAL officer, to discuss cross-country patterns and final recommendations.

4 From practice to insight: Key lessons from systems evaluation in WASH

Based on the application of the evaluation approach in Welthungerhilfe's Global WASH Programme, several key lessons have emerged. These insights are not only relevant for Welthungerhilfe's Global WASH Program but also offer practical guidance for other WASH systems strengthening programmes and systems-focused interventions in related sectors.

A clear results chain is foundational

One of the most critical prerequisites for evaluating a WASH systems strengthening programme is the existence of a clearly articulated results chain. This should define the specific interventions to be undertaken - at different system levels - and outline how these are expected to lead to outputs, intermediate outcomes, and ultimately impact. In programmes of limited duration (e.g. 3–4 years), a well-structured results chain allows for the identification of realistic and measurable intermediate outcomes, while also clarifying the logic of how change is expected to happen. It serves not only as a planning and monitoring tool, but also enables evaluators to assess whether the programme is functioning as intended, and whether teams understand the system and their role in influencing it.

Reflection and adaptation are integral to evaluation

Systems strengthening is inherently non-linear and dynamic. Regular reflection and review - embedded within programme implementation - should not only serve as internal checks, but also as drivers of progress and sources of evaluative evidence. Structured reflection moments (e.g. learning workshops, adaptive planning sessions) and their documentation can capture how and why programme strategies evolved, and whether these adaptations increased relevance and coherence. For evaluators, these internal processes provide essential context for interpreting observed outcomes and understanding the programme's responsiveness to complexity.

Process evaluation early on, effect evaluation later

The timing of evaluation activities matters. In the early phases of a WASH systems programme, process evaluation is particularly important to assess whether diagnostic work was carried out effectively, whether the intervention logic is well-aligned with context, and whether stakeholder coordination has been prioritised. Midterm reviews are well placed to assess whether the programme is on the right track and whether the results chain remains valid. Given that the effects of systems interventions often take time to emerge, effect evaluation is best conducted after several years, when there is a greater likelihood of observing changes in system behaviour, institutional performance, or coordination mechanisms.

Qualitative tools are needed to capture systemic outcomes

By nature, WASH systems strengthening aims to influence how actors within the system fulfil their roles - e.g. through better planning, coordination, regulation, or accountability. These changes are often qualitative and distributed across actors and levels. While interviews and field observations are helpful, they are insufficient on their own. Tools like outcome harvesting and building block assessments can support systematic, evidence-based monitoring of intermediate outcomes. These tools require programme teams to substantiate claims of change with concrete examples and justifications, which strengthens both internal learning and external evaluation.

Systemic impact requires a broader definition

Welthungerhilfe's updated definition of impact (refer to Annex 2) aligns well with the OECD's revised evaluation criteria and provides a more appropriate lens for systems evaluations. Rather than focusing solely on individual-level well-being (e.g. health outcomes), systemic impact is understood as enduring changes in patterns, structures, and paradigms - such as the establishment of a regulatory authority or the adoption of decentralised service models. Given the difficulty of attributing downstream human development outcomes to a single systems intervention, it is more meaningful for WASH systems programmes to define impact in terms of changes to the system itself. This reframing enables evaluations to capture what systems strengthening is truly about: transforming how services are governed, financed, and delivered over time.

References

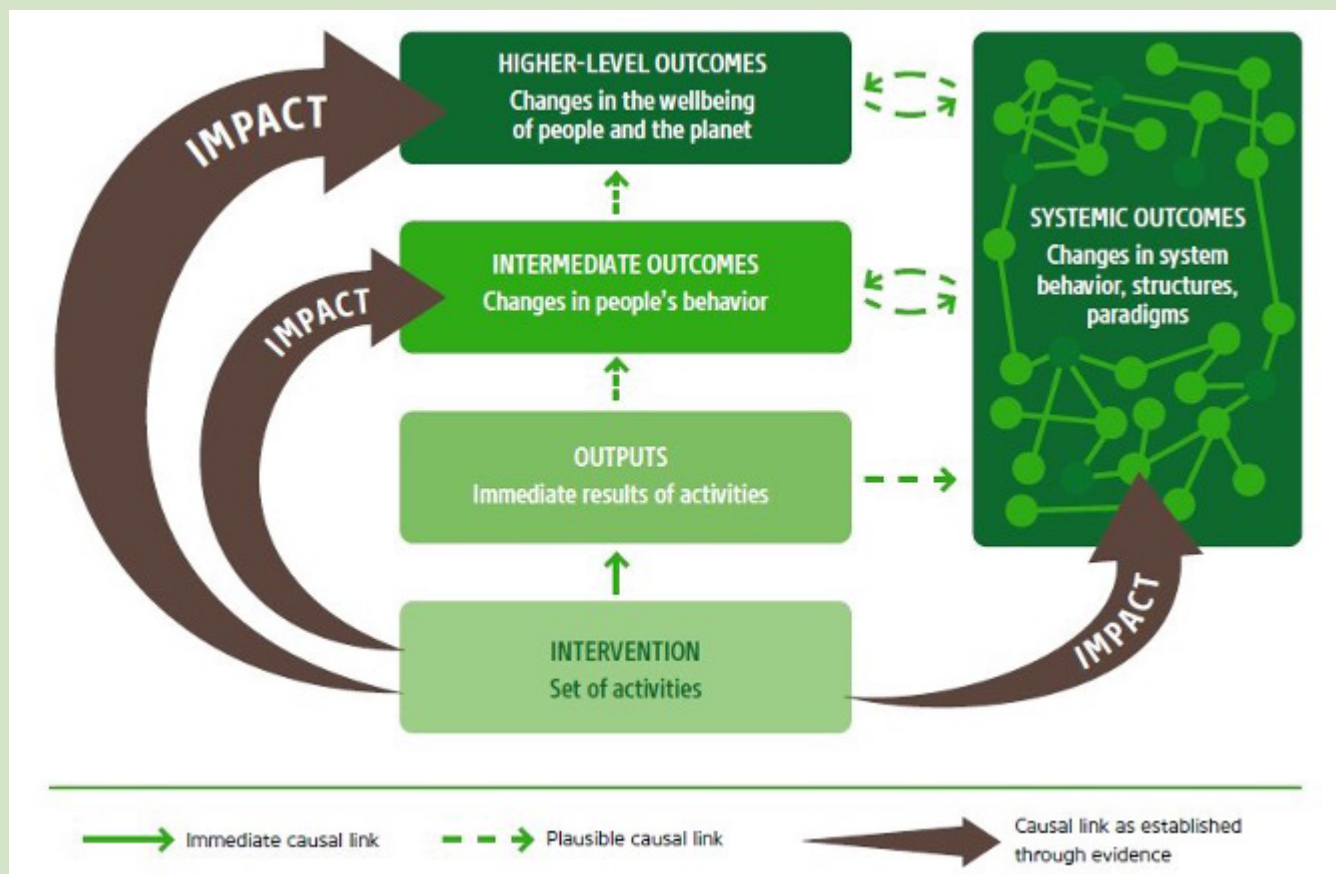
- AMCOW (African Ministers Council on Water) (2011). AMCOW Country Status Overviews – Regional synthesis report. Pathways to progress: Transitioning to country-led service delivery pathways to meet Africa's water supply and sanitation targets
- Battle, C., 2019. Beyond building blocks?: identifying and monitoring dynamic drivers of sector performance: policy brief. London, UK: WaterAid. Available at: <https://washmatters.wateraid.org/sites/g/files/jkxoof256/files/beyond-building-blocks-identifying-and-monitoring-drivers-of-sector-performance-policy-brief.pdf>
- Casey, V. & Crichton-Smith, H., 2020. System strengthening for inclusive, lasting WASH that transforms people's lives : practical experiences from the SusWASH programme. London, UK: WaterAid. <https://washmatters.wateraid.org/publications/suswash-system-strengthening-for-inclusive-lasting-water-sanitation-hygiene>
- Huston, A. & Moriarty, P.B., 2018. Understanding the WASH system and its building blocks : building strong WASH systems for the SDGs, (IRC WASH systems series) The Hague, the Netherlands: IRC. Available at: https://www.ircwash.org/sites/default/files/084-201813wp_buildingblocksdef_newweb.pdf
- OECD. 2019. Better Criteria for Better Evaluation; Revised Evaluation Criteria, Definitions and Principles for Use. OECD/ DAC Network on Development Evaluation. <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>
- Smits, S. and J. Butterworth. 2023. Evaluative research of the Global Program on WASH Systems Strengthening; the proposed approach. Welthungerhilfe, Bonn, Germany
- Smits, S. and R. Shiva. 2024. Mid-Term Review of the Global Program on WASH Systems Strengthening; synthesis report. Welthungerhilfe, Bonn, Germany
- Smits, S. and R. Shiva. 2025. Evaluation of the Global Program on WASH Systems Strengthening; synthesis report. Welthungerhilfe, Bonn, Germany
- Stanford Program on Water, Health & Development, 2020. Conrad N. Hilton Foundation Safe Water Initiative: strategy measurement, evaluation & learning : selecting comparator districts. Stanford, CA, USA: Stanford Woods Institute for the Environment.
- UNICEF (2018). Water, sanitation and hygiene bottleneck analysis tool (WASH BAT) user guide. Available at www.washbat.org/wp-content/uploads/en/WASH_BAT_2_User_Manual.pdf
- Valcourt, N., Javernick-Wil, A., Walters, J., & Linden K. 2020. System approaches to water, sanitation, and hygiene : a systematic literature review. International journal of environmental research and public health, 17(3), 702. DOI: [10.3390/ijerph17030702](https://doi.org/10.3390/ijerph17030702)
- WHH. 2024. Impact Paper Series; Episode 1.1 Defining impact.
- WHO. 2010. Monitoring the building blocks of health systems : a handbook of indicators and their measurement strategies. Geneva, Switzerland: World Health Organization. Available at: [9789241564052-eng.pdf \(who.int\)](https://www.who.int/publications/i/item/9789241564052-eng)

Annex 1: Detailed evaluation questions

Research area	Evaluation questions of the MTR	Evaluation questions of the end evaluation
1. Relevance of systems strengthening	<ul style="list-style-type: none"> To what extent has the current system in the focus districts enabled progress against WASH targets up till now, and is it projected to achieve those without any additional system interventions? What is the specific relevance of undertaking systems interventions? 	<ul style="list-style-type: none"> Not included in the end evaluation
2a. Relevance and coherence of the adopted intervention logic	<ul style="list-style-type: none"> To what extent does the intervention logic address the specific systemic limitations that are currently holding back the achievement of WASH targets in the focus districts? To what extent is the intervention logic aligned with, and complementary to, other systems strengthening efforts in the sector and in the focus districts? Is the intervention logic clear and internally consistent? To what extent has GWP adapted the intervention logic to changing insights and contextual changes in the course of the programme? 	<ul style="list-style-type: none"> To what extent has the programme adapted the intervention logic to changing insights and contextual changes since the completion of the mid-term review? To what extent has the GWP triggered other stakeholders (local and national government, private sector, other development partners) to undertake complementary efforts towards strengthening WASH systems and/or found complementarity with efforts of those partners?
2b. Coherence of interventions with likely system-wide ability to sustain them	<ul style="list-style-type: none"> To what extent are the interventions in line with the likely system-wide ability to sustain these? To what extent can the interventions be embedded or included in government structures and processes? To what extent are other actors creating further enabling conditions? 	<ul style="list-style-type: none"> To what extent are the interventions in line with the likely system-wide ability to sustain these? To what extent can the interventions be embedded or included in governments structures and processes? To what extent are other actors creating further enabling conditions?
3. Effectiveness of the process	<p>Programme level:</p> <ul style="list-style-type: none"> To what extent is the set-up, organisation, capacity, and methodology of the GWP (potentially) effective? To what extent have methodologies and tools been provided to support GWP countries in establishing the process <p>Country level:</p> <ul style="list-style-type: none"> Which of the process steps – with corresponding methodology and tools – that were undertaken were necessary to define, implement and adjust the systems interventions? How effective have the capacity development activities for GWP country staff and partners been? 	<ul style="list-style-type: none"> To what extent have adjustments been made to the set-up, organisation and methodology of the GWP, both at programme and country level? How effective have these been?
4. Effectiveness and efficiency of results	<ul style="list-style-type: none"> Are the systems and processes in place to track the achievement of outputs and contributions to outcomes and impacts? Which of the intended outputs of the interventions were achieved to date? Which of the intended immediate and intermediate outcomes (as per the results chain) did the interventions contribute to? Were there additional intermediate outcomes that the interventions contributed to? Is budget expenditure on track? 	<ul style="list-style-type: none"> To what extent have adjustments been made to the set-up, organisation and methodology of the GWP, both at programme and country level? How effective have these been? What changes have been made to the systems and processes in place to track the achievement of outputs and contribution to outcomes and impacts? Which of the intended outputs of the interventions have been achieved? Which of the intended intermediate outcomes (as per the results chain) did the interventions contribute to? Were there additional intermediate outcomes that the interventions contributed to? Are there any intermediate outcomes related to other actors being triggered or enabled to undertake efforts towards system strengthening? And can these be quantified? Is budget expenditure on track?
5. Impact	<ul style="list-style-type: none"> What do the patterns of change evident in the outcomes (expected as well as unexpected) suggest about the WASH system trajectory? What did the project contribute to these outcomes? And what was the current trajectory of the WASH sector? What would changes in the WASH system trajectory imply for sustainable, equitable realization of WASH services? 	<ul style="list-style-type: none"> What do the patterns of change evident in the outcomes (expected as well as unexpected) suggest about the WASH system trajectory? What did the project contribute to these outcomes? And what was the already ongoing trajectory of the WASH sector? What would change in the WASH system trajectory imply with respect to sustainable, equitable realization of the WASH services?

Annex 2: Welthungerhilfe's Outcome and Impact Framework

Welthungerhilfe has updated its terminology for outcomes and impact to better reflect the complexity of change in systems strengthening contexts. The paper [Defining Impact](#) (2024) introduces a practical frame-work for capturing systemic impact—shifting focus from individual-level outcomes to enduring changes in patterns, structures, and paradigms. The approach which is illustrated in the diagram below is especially relevant for WASH, food systems, and other multi-actor development sectors.



The diagram highlights the importance of causality (contribution or attribution) between interventions and different types of outcomes. The WHH framework gives examples of different types of impact associated with WASH interventions:

WASH Deliverables	Change in WASH behaviour	Change in wellbeing	Change in patterns, structures, paradigms
Project participants are trained on water purification and hygiene practices. Participating households receive essential WASH supplies, such as water filters and soaps.	Project participants start using basic sanitation services and basic drinking water services.	Decrease in water-related illnesses	<p>Pattern change: Government officials become more responsive to people's complaints.</p> <p>Structure change: establishment of a regulatory authority responsible for overseeing and regulating WASH services</p> <p>Paradigm change: Transition from centralized waste-water treatment to decentralized and nature-based solutions, promoting ecological and sustainable approaches.</p>

Contact

mailto: wash@welthungerhilfe.de
www.welthungerhilfe.org
www.welthungerhilfe.org/wash
 mailto: info@ircwash.org