

Evaluating & Improving the WASH sector

Strengthening WASH Governance

Learning about Complexity

Assessing Change

Thematic Overview Paper 23

Main author: Régis Garandau

Contributing authors: Deirdre Casella and Kristof Bostoën (IRC)

Reviewed by: Nadia Manning-Thomas (IWMI), Patricia Rogers (RMIT)
and Howard White (3ie)



Evaluating & Improving the WASH Sector

**Strengthening WASH Governance
Learning about Complexity
Assessing Change**

Thematic Overview Paper 23

Main author: Régis Garandeau

Contributing authors: Deirdre Casella and Kristof Bostoen (IRC)

Reviewed by: Nadia Manning-Thomas (IWMI), Patricia Rogers (RMIT)
and Howard White (3ie)

Thematic Overview Papers (TOPs) are a web-based series. However, those who don't have access to the Internet should also be able to benefit from them. This is why we make paper versions of TOPs available as well as putting them on line.

Edited by: Peter McIntyre

Copyright © IRC International Water and Sanitation Centre (2009)

IRC enjoys copyright under Protocol 2 of the Universal Copyright Convention.

Nevertheless, permission is hereby granted for reproduction of this material, in whole or in part, for educational, scientific, or development related purposes except those involving commercial sale, provided that (a) full citation of the source is given and (b) notification is given in writing to IRC, P.O. Box 82327, 2508 EH The Hague, The Netherlands, Tel: +31 (0)70 3044000, Fax: +31 (0)70 3044044, e-mail: publications@irc.nl

Table of Content

Thematic Overview Papers (TOPs): an effective way to TOP up your knowledge	3
Contents of each TOP	3
1. Introduction: Three paradigms of development and its evaluation	4
1.1 A tale of three farms	4
1.2 ...Or three paradigms of development interventions and their evaluation	6
1.3 Why and for whom are we measuring development's successes and failures?	8
2. The first paradigm: The causal chain and evidence-based evaluations	10
2.1 The causal chain of development interventions, from inputs to impacts	10
2.2 Longitudinal studies	13
2.3 Case-control studies	14
2.4 Experimental evaluations, or Randomised Controlled Trials	16
2.5 Quasi-experimental evaluations	18
2.6 Non-experimental evaluations	19
3. The second paradigm: Black boxes and realistic evaluations	22
3.1 The black boxes of development interventions and the theories to open them	22
3.2 The Minimum Evaluation Procedures	24
3.3 The Hygiene Evaluation Procedures	26
3.4 From impacts to outcomes, context and processes, from quantitative to qualitative analysis	26
4. The third paradigm: Complex adaptive systems, complexity evaluations and change assessment	28
4.1 The WASH sector and its increasing complexity	28
4.2 The WASH sector and the evaluation gaps	30
4.3 The third paradigm: complex systems, complexity evaluations	33
4.4 Why and for whom should we measure the WASH sector's successes and failures?	35
4.5 Evaluation or assessment?	36
5. Some examples of participatory tools to assess change in complex WASH interventions	38
5.1 Balanced Score Card	38
5.2 Most Significant Change	41
5.3 Social Network Analysis	43
5.4 Outcome mapping	45
5.5 Process documentation	47
5.6 Qualitative information system	49
6. Conclusion and the way forward: Mixed evaluation and assessment methods	51

6.1	Conclusion.....	51
6.2	Summary table of all the evaluation and assessment methods above.....	53
7.	Glossary and acronyms.....	59
7.1	Acronyms	59
7.2	Glossary	60
8.	TOP books, articles, papers	73
8.1	Evidence-based, theory-based and realistic evaluations.....	73
8.2	Water Governance	73
8.3	Complexity and development.....	73
8.4	Qualitative assessment tools	74
9.	TOP websites	76
9.1	International Evaluation Networks	76
9.2	Regional Evaluation Associations	77
9.3	Development Banks Evaluation Departments	78
9.4	Research organisations on monitoring, evaluation, assessment and development.....	79
9.5	Other Monitoring, evaluation and assessment websites	80
10.	TOP contacts	82
11.	References	86
12.	About IRC	98

Thematic Overview Papers (TOPs): an effective way to TOP up your knowledge

Do you need to get up to speed quickly on current thinking about a critical issue in the field of water, sanitation and health?

Try an IRC TOP (Thematic Overview Paper).

TOPs are a web-based initiative from IRC. They combine a concise digest of recent experiences, expert opinions and foreseeable trends with links to the most informative publications, websites and research information. Each TOP will contain enough immediate information to give a grounding in the topic concerned, with direct access to more detailed coverage of your own special interests, plus contact details of resource centres or individuals who can give local help.

Reviewed by recognised experts and updated continually with new case studies, research findings, etc, TOPs will provide water, sanitation and health professionals with a single source of the most up-to-date thinking and knowledge in the sector.

Contents of each TOP

Each TOP consists of:

- An Overview Paper with all the latest thinking
- Case studies of best practice, if applicable
- TOP Resources:
 - links to books, papers, articles
 - links to web sites with additional information
 - links to contact details for resource centres, information networks or individual experts
 - a chance to feedback your own experiences or to ask questions via the Web.

The website will contain a pdf version of the most up-to-date version of the TOP, so that individuals can download and print the information to share with colleagues.

TOPs are intended as dossiers to meet the needs of water, sanitation and health professionals in the South and the North, working for national and local government, NGOs, community-based organisations, resource centres, private sector firms, UN agencies and multilateral or bilateral support agencies.

1. Introduction: Three paradigms of development and its evaluation

“What we achieve inwardly
will change outer reality.”

Plutarch

1.1 A tale of three farms...

Once upon a time, a young boy moved to a new region to find some work. Upon arrival, he asked people where he could get employed and heard:

- “There are three farms around. You could ask the farmers if they need a hand.”

The young boy went to the first farm and asked if he could work for the farmer.

- “What do you know about farming?” he asked.

- “Not much, Sir.”

- “This is fine, because farming is relatively simple. We decide how much and what kind of vegetables we want each season, then we calculate how many seeds, how many tools, and how many workers we need and we define a farming strategy. Then we cultivate plots of land, and grow and harvest the vegetables. Then we simply count the quantity of each type of vegetables we harvest, and see if the farming strategy has worked or not. If it has not succeeded, we improve it in the next season, so that we can produce more vegetables.”

- “That sounds good to me.”

And the boy started to work on the farm the next day. Together with the farmer and the other workers, they decided how many of each type of vegetable they wanted, and made their plans accordingly. They cultivated plots of lands, harvested and counted the vegetables, and tried to improve in the next season. But after four seasons, he realised they never reached their objectives: there was always too little of certain varieties of vegetables, sometimes too much of another, and they could not always harvest everything properly and in time with the tools they had.

- “Our objectives are not realistic” he told the other workers, “we work hard, but production varies a lot from plot to plot, and also depends on the weather. We try to improve, but it never works out as planned. We need to change the way we work. And the tools we have are not always adapted to the conditions; we need to change them as well. We need to take more notice of changes in the soil and the weather.”

- “We can repair the tools and continue to try and improve our plans and strategies,” said the other workers. “But we cannot waste our time worrying about the weather, otherwise the farmer will think we are lazy!”

- “What a pity” thought the young boy, “We are doing well but we could learn so much more.”

And he left the first farm to try and find another job.

The young boy then went to the second farm, and proposed his work to the farmer.

- "What do you know about farming?" he asked.

- "A little, Sir, I have already worked four seasons on a farm."

- "Good, because farming is quite complicated: we decide how much and what kind of vegetables we want for each season, then we calculate how many seeds, how many tools, and how many workers we need and we define a farming strategy. Then we cultivate plots of land, and grow and harvest the vegetables. Then, we count the quantity of each type of vegetables we got. But this is where it gets complicated, because we have to take into account soil conditions in each plot and the weather. We see if the farming strategy has worked or not in this context. If not, we can improve it in the next season, and adapt to the conditions so that we can produce more and better vegetables."

- "That sounds good to me."

And the boy started to work in the farm the next day. Together with the farmer and the other workers, they decided how many of each type of vegetable they wanted, and made their plans accordingly. They cultivated plots of land, harvested and counted the vegetables, tried to understand which farming strategy worked best on which land and with which weather and tried to improve in the next season. But after four seasons, he realised they never reached their objectives: there was always too little of certain varieties of vegetables, sometimes too much of another, and they could not always harvest everything properly and in time with the tools they had.

- "Our objectives are not realistic" he told the other workers, "we work hard and try to predict and adapt to the context, but production varies a lot, not only depending on the land and weather. It also depends on when the chief of the village allows us to plant seed and to harvest, and whether the seed seller delivers on time or not. When we harvest the vegetables, what we manage to sell depends on what villagers want to buy, and how much other farms produced. It depends on so many other things and so many other people! We try to improve, but it never works out as planned. We need to change the way we work. And the tools we have to harvest are not always adapted, we need to change them as well. We need to take many more things into account before we can plan for success."

- "We can repair the tools and continue to try and improve our plans and strategies," said the other workers. "But we cannot start worrying about what all these other people are doing, otherwise the farmer will think we are lazy!"

- "What a pity" thought the young boy, "We are doing well but we could learn so much more."

And he left the second farm to try and find another job.

He finally went to the third farm and proposed his work to the farmer.

- "What do you know about farming?" he asked.

- "Quite a lot, Sir, I have already worked eight seasons on two different farms. But there are still so many things I do not understand. We defined what and how much we wanted to

produce, tried different farming techniques and checked which one worked and which one did not, but it was not sufficient. We tried to check where and when various farming techniques worked and when they did not. But it is still not enough. Each time there are other things and people who influence our production and change our plans.”

- “This is normal,” he answered. “Farming is complex: you can never predict how many vegetables of each type you will harvest and sell as it depends on so many factors, some that you can foresee and control, some that you can only influence and others you cannot do anything about. The most complex and unpredictable factor is people. All of these factors can change, and interact together, in a way that is still too complex to understand. Counting the vegetables and trying to improve and adapt your farming techniques does help. But trying to find what works and what does not, where and when is not always enough. Sometimes, you also need to find out how, why and with whom it works or not. You need to understand what processes and what changes can improve your harvest.”

- “And how can I achieve this?”

- “First, you need to change yourself, the way you think and the way you work. Be more humble and more patient. Then act, count, but also listen, observe. And learn.”

1.2 ...Or three paradigms of development interventions and their evaluation

Vegetables certainly are tasty and healthy food, but what exactly does this short tale tell us, apart from the fact that farming is more complex than it appears?

It describes three farms with different ways of conceptualising farming and the way it can be evaluated and improved: a relatively simple causal one, a more complicated but still causal one, and a complex one that recognises the differences in many variables in addition to the crops and conditions, some of them highly unpredictable, perhaps due to the fact their causal relationships are not fully understood. Let us now replace the farmers by development funding agencies, the workers by implementing agencies, and the farming seasons by cycles of development interventions. The farming techniques, or intervention designs, aim at producing the desired amounts of vegetables, i.e. impacts. These are harvested with different tools, or evaluation methods. There are sometimes too many of some varieties and too little of others, which highlights there can be over- or under-estimations of some impacts, maybe because of imperfect evaluation tools or because certain types of impact are given more attention than others. Production varies depending on the plots of land and weather, i.e. the geographical and environmental contexts and the social and cultural aspects that go with them. But it also depends on the seed dealer and the village chief, representing the intervention partners and local authorities, or generally speaking institutions. And finally, the villagers who buy, or do not buy, the vegetables, are the beneficiaries. This tale of three farms then becomes a metaphor for three different – but complementary – conceptualisations, or paradigms, of how development interventions work, and how they can be evaluated and improved.

As in the first farm, the first paradigm sees development interventions as a static, single chain of events resulting from one another according to a precise and previously established plan of causes, effects and assumptions ending with predictable impacts. This paradigm leads development agencies to evaluate these impacts quantitatively, with little concern for the causal chain leading to them as it is supposed to follow the initial plan. Whether the impacts measured at the end are as numerous and effective as was initially planned will determine whether the original plan was right or not.

As in the second farm, the second paradigm still sees development interventions as planned chains of events linked by cause and effect relationships and leading to predictable impacts, but recognises that the context influences these mechanisms and might create multiple or alternative strands in causal chains. In various contexts, different causes and effects mechanisms can lead to similar impacts. This paradigm leads development agencies to evaluate these impacts taking the contexts and mechanisms in which they are produced into account. The quantity of impacts measured in different contexts, and through different mechanisms, will then determine in which contexts and through which mechanisms the initial plan can work or not.

Both these paradigms are useful to implement and evaluate certain types of development interventions, particularly discrete, standardised interventions such those aiming to build physical infrastructure or change behaviour of individuals, e.g. installation of handpumps or handwashing promotion. But some argue that they are less adapted to evaluate interventions aiming at bringing a complex mixture of social, institutional and political change, i.e. software and governance issues, at the household, community and local governments levels. Numerous technical, environmental, but also social, political, institutional, financial and economic factors influence development interventions, their impacts and their sustainability. All these factors are interrelated and influence each other, constantly changing the context in which an intervention is implemented, and the mechanisms through which it can be implemented. This dynamic social and political environment is sometimes too complex to be conceptualised only in terms of linear (or simple non-linear) and uni-directional causes and effects; it has unpredictable consequences on complex development interventions and the processes through which they might, or might not, change things and produce impacts that could be quantified in a meaningful way. Impacts of complex interventions are therefore rather unpredictable themselves, and often difficult to attribute to a specific intervention, which has to be taken into account with appropriate evaluation and assessment tools.

As in the case of the third farm, there is therefore a need to take complexity into greater account, to adapt the previous paradigms of development interventions and the resulting evaluation methods. Evaluating what works and what does not, where and when is not always enough. Complementary qualitative tools already exist to assess dynamic social and institutional systems, their changes and the processes leading to impacts. These will be presented in due time but, to start with, let us come back to the first two farms and their paradigms of interventions and evaluation.

1.3 Why and for whom are we measuring development's successes and failures?

All development agencies try to measure in one way or another the performance and achievements of their interventions. The need for such measurements mostly comes from two traits common to all interventions and agencies:

- Development interventions are planned in the form of projects, which usually have specific, initially agreed targets and goals within a determined time frame. Implementers therefore need to measure the progress of their work and possible deviations from the initial plan, to be able to complete the promised outputs in a timely fashion in the right quantity and quality, as well as to report regularly on their activities: this first type of measurement is *monitoring*¹.
- Funding agencies but sometimes also other intervention partners, request that the implementers should be transparent and *accountable*, i.e. demonstrate that resources are used accurately and produce the desired impacts, hence the need for measuring the efficiency of an intervention, and for quantifying its impacts: this second type of measurement is *evaluation*. Some prefer the term assessment, although it is less commonly used. Some differences between evaluation and assessment will be discussed later in the document (see 4.5.).

Most development agencies therefore have a Monitoring and Evaluation (M&E) department or unit, although each might have a different understanding of these terms. The broadest definition actually unifies both terms, describing *comprehensive evaluation* as including needs analysis, ex-ante impact evaluation, monitoring, process evaluation, cost-benefit evaluation, and impact evaluation. Monitoring helps to assess whether an intervention is being implemented as was planned, giving continuous feedback on the status of implementation and identifying specific problems as they arise. *Process evaluation* (also sometimes called *formative evaluation*) is concerned with how the intervention operates and focuses on problems in service delivery. *Cost-benefit analysis* and *cost-effectiveness evaluations* assess programme costs (monetary or non monetary), in relation to the benefits being produced by the intervention, and in comparison to alternative uses of the same resources. And finally, *impact evaluation* (also sometimes called *summative evaluation*) is intended to determine more broadly whether the intervention had the desired effects on individuals, households, and institutions and whether those effects are attributable to the intervention. Impact evaluations can also explore unintended, positive or negative consequences on *beneficiaries* (after Baker, 2000).

To measure progress, deviation, operation costs and impacts, monitoring and evaluation rely on *indicators*, sometimes qualitative but mostly quantitative. Indicators are simple numbers, descriptive or normative statements that can condense complex information on

¹ Terms in italics are defined in the [Glossary](#) at the end of the TOP.

systems and processes to a manageable amount of meaningful information. They can be used to observe, describe, and evaluate actual states, to formulate desired states or to compare an actual with a desired state (UNWWAP, 2003, quoted in WMA, 2005).

Over recent decades, several events and trends have pointed to an increased role for monitoring and evaluation in development interventions. World development roadmaps like the International Drinking Water Supply and Sanitation Decade (IDWSSD, 1981-1990) or the Millennium Development Goals (MDGs, 1990) set quantitative development targets to be achieved by countries and development agencies to tackle poverty, therefore increasing the need to measure development performance, both at the intervention and agency levels, but also at the national and global ones. This led to the creation of the Joint Monitoring Programme (JMP), started in 1990 by the World Health Organisation (WHO) and the United Nations International Children's Fund (UNICEF) to monitor water and sanitation coverage. The introduction of results-based management (RBM) into various development agencies in the mid-1990s also reinforced the focus on development performance and impacts (IEG, 2008). As part of this trend, funding agencies started to request more objective, independent evaluations to prove that their resources are used transparently on interventions that improve people's lives, health, education, and wealth. All these events and trends resulted in an increased focus on development agencies' results, hence an increased demand for quantitative measurement in general, and impact evaluation in particular.

This TOP will therefore mostly discuss evaluations, focusing on the Water, Sanitation and Hygiene (WASH) sector. Each of the three paradigms of development will be discussed separately in the specific context of this sector. For each paradigm, the resulting evaluation methodologies will be presented in a chronological order, illustrated by examples of WASH related evaluations.

2. The first paradigm: The causal chain and evidence-based evaluations

“A theory can be proved by experiment,
but no path leads from experiment to the birth of a theory.”

Albert Einstein

2.1 The causal chain of development interventions, from inputs to impacts

The first paradigm of development interventions pictures them as series of steps linking an intervention's inputs to its intended results through direct cause and effect relationships. These are therefore called the *causal chain* (or the *results chain*), which is presented in Figure 1 and detailed below.

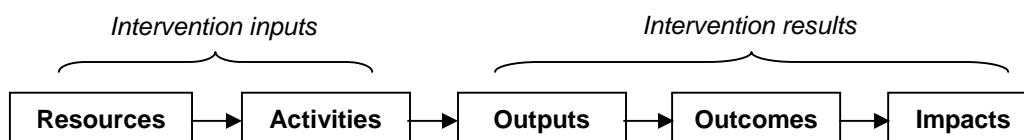


Figure 1. the causal chain of development interventions (after Poulos et al., 2006)

Resources are the financial, human, social, institutional and material resources supporting or constraining the intervention. Some are external to the intervention (e.g. legal or regulatory environment; institutions, their staff and skills; communities, their members and culture), while others are put in by the intervention (Poulos et al., 2006) (e.g. budget, project staff, material and equipment).

Activities are the actions and processes planned and implemented on the basis of the resources, as described in the intervention design or corrected through monitoring and evaluation (Poulos et al., 2006) (e.g. construction of infrastructure with material, training of staff or beneficiaries, system upgrading).

Outputs are the products, capital, goods and services which result directly from the intervention activities (Poulos et al., 2006) (e.g. infrastructure built, people trained, or upgraded system).

Outcomes are the short or medium term changes in relationships, activities, actions, knowledge or behaviours of the intervention partners (institutions and individuals, including beneficiaries) that can be linked to its activities although they are not necessarily directly caused by it (after Earl et al., 2001a).

Impacts are the ultimate, positive and negative, intended or unintended, primary and secondary long term effects produced by a development intervention, directly or indirectly,

on beneficiaries or other people (DAC, 2002) (e.g. better and more sustainable access to infrastructure and services, time-saving, improved living conditions).

Finally, the *causal path* is the way the intervention moves from one link to another, in other words, the arrows between the links of the causal chain. Some interventions might fail because the implementation of the causal chain failed, or because the logics behind the causal chain were wrong.

This causal chain actually is a simplified version of the *logical framework*, often used to design and summarise development interventions and their monitoring and evaluation systems, including indicators and the way to measure them. Table 1 (following page) shows a typical logical framework matrix, and its links to the causal chain.

This paradigm of development interventions as planned chains of steps eventually causing predictable impacts led development agencies to design evaluations to quantify these impacts and attribute them (or establish a causal link) to the interventions and nothing else (Figure 2).

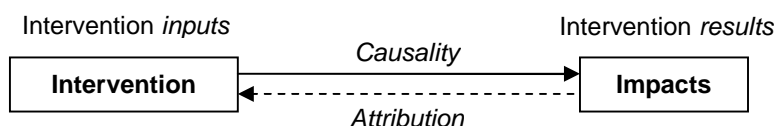


Figure 2. the causality/attribution links between an intervention and its impacts

When planning and implementing WASH interventions, one of the main objectives of development agencies is usually to produce positive impacts on people's health (Figure 3a). Agencies therefore regularly attempted to evaluate the health impacts that could be attributed to their WASH interventions (Figure 3b). However, if water, sanitation and hygiene behaviours often have a considerable impact on health, many other factors also contribute to people's health status (Figure 3c). These various factors and their complex interaction make it difficult to unambiguously relate health changes to WASH interventions alone (Figure 3d) (Bostoën, 2007).

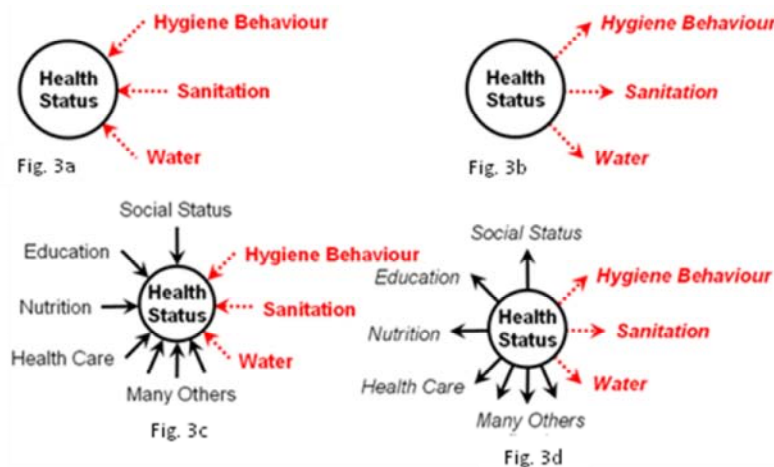


Figure 3 (a, b, c, d). Relation between water, sanitation, hygiene and health (Bostoën, 2007)

Table 1. The logical framework matrix and its links with monitoring and evaluation (after NZAid, 2007 and White, 2006)

	Causal chain (Narrative summary)	Objectively Verifiable Indicators	Means of verification	External factors & assumptions
Intervention results	Impacts (<i>goals</i>) Desired final development impacts to which activities should contribute	Quantitative or qualitative evaluation indicators measuring progress towards desired the impacts	How the information will be collected, when and by whom and how it will be reported.	Factors outside the intervention control needed to sustain impacts in the long term. Assumption that the causal chain operates smoothly.
	Outcomes (<i>objectives or purpose</i>) Desired changes to which activities should contribute	Quantitative or qualitative evaluation indicators measuring changes and progress towards the desired outcomes	As above.	Factors outside the intervention control that may impact on the 'outcomes to impacts' link. Assumption that the causal chain operates smoothly.
	Outputs The products and/or services delivered by the activities which are under the implementation management's control.	Quantitative or qualitative monitoring indicators measuring progress towards delivering the planned outcomes	As above.	Factors outside the intervention control that may impact on the 'outputs to outcomes' link. Assumption that the causal chain operates smoothly.
Intervention inputs	Activities The tasks that have to be completed to deliver the planned outputs.	Quantitative or qualitative monitoring indicators measuring progress towards implementing the planned activities	As above.	Factors outside the intervention control that may impact on the 'activities to outputs' link. Assumption that the causal chain operates smoothly.
	Resources (usually not included in the logical framework matrix)	Quantitative or qualitative monitoring indicators measuring progress towards gathering and using the necessary external or internal resources		

This difficulty led agencies to develop more and more sophisticated and ‘rigorous’ *evidence-based evaluations*, using various evaluation methodologies, to try to quantify health improvements precisely, and to give scientific evidence that these were caused by specific WASH interventions. These methods are largely inspired by epidemiological and medical research and trials, and the related jargon (including the term ‘evidence-based’). They include longitudinal and case-control studies, as well as experimental, quasi-experimental and non-experimental impact evaluations, which are described below.

2.2 Longitudinal studies

Until the mid-1970s, the main tool used to evaluate the health impacts of WASH interventions were large scale *longitudinal studies* (Cairncross, 1999). These basically consist of comparing the progress of health indicators measured within the same group of beneficiaries before and after the intervention. For instance, a longitudinal study could compare the incidence of diarrhoea in a region before and after a large WASH intervention to provide the communities of this region with wells, boreholes and latrines, to quantify the health impact of these infrastructures. Long term longitudinal studies can even repeat observations on the same group of people for years, or even decade.

Such *reflexive comparisons* (i.e. before/after) are useful when evaluating a full coverage or country-wide intervention where beneficiaries cannot be compared to others within the same region (World Bank, 2009). This methodology was for instance used to evaluate the impact of a city-wide sanitation programme on diarrhoea, in Brazil (Barreto et al., 2007). But their main weakness is that they do not take into account how the health indicators of this specific population would have evolved without the intervention, i.e. the *counterfactual* (Ravallion, 2008). Part of the variation of the health indicators measured before and after the intervention might be due to unaccounted for external factors influencing people’s lives, but not caused by the intervention, for instance, a general trend towards better health due to improving economic conditions, or towards worse health in case of a famine (factors). The failure to take such factors, called *confounders* (or *confounding factors*), into account, can call the results of health impact studies into question (Cairncross et al., 1997). They make it difficult for longitudinal studies to quantify and attribute with certitude the measured health changes to WASH interventions (Figure 4). Longitudinal studies can still be useful to identify or rule out some hypotheses and partly explain what is happening and why, but

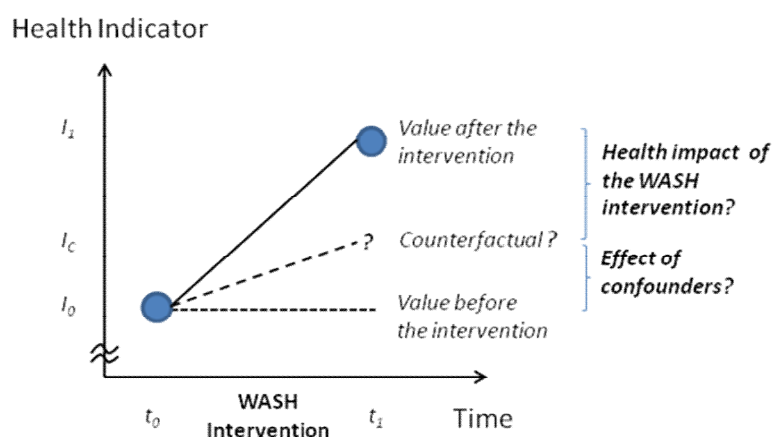


Figure 4. Longitudinal studies, the counterfactual and confounders

other, complementary methods often have to be used to further investigate their results.

In spite of the huge efforts and vast amounts of money put into such studies, it soon became clear that they were not most suitable for the WASH sector, as they did not show the real health impacts of WASH interventions, but more general health trends. This lack of meaningful results led a panel of experts to conclude that the World Bank should not undertake or invest in any such long-term longitudinal studies (World Bank, 1976). A review of the work of this panel even suggested that trying to quantify the specific health impacts of water interventions is futile, as governmental, physical, environmental, economic, cultural and educational factors have significant impacts upon disease and health, and isolating the water factor alone is too difficult (Wolman, 1976).

2.3 Case-control studies

With the development of new epidemiological methods, new attempts were made to measure the health impacts of WASH interventions. During the 1980s and the International Drinking Water Supply and Sanitation Decade, *case-control studies* were seen as a quick and cheap way to retrospectively prove and quantify the links between people's health status and their access to WASH infrastructure (Briscoe et al., 1985, quoted in Cairncross, 1999).

Case-control studies compare two groups of people, one with health symptoms such as diarrhoea (*cases*) and one without (*controls*), and look at their access to WASH infrastructure to try and determine whether those with better access are less likely to present the symptoms (Hunter et al., 2002). For instance, Young et al. (1987) conducted such a study in Malawi and determined that children living in families who use good quality water supplies and latrines experience 20% less diarrhoea than those who do not.

However, case-control studies might not give meaningful results unless they ensure that the two groups are on average identical, for instance in socio-economic characteristics (wealth, education, etc.), as differences in these might explain part of the difference in their respective measured indicators. This is called *selection bias* (Ravallion, 2008). For example, a case-control study might establish a link between private wells, latrines or even television ownership, and reduced incidence of diarrhoea. However, households who own a private well, a latrine or a television are likely to be wealthier and better-educated compared to those who do not, which probably contributes to protecting them from illness, causing an apparent association which is not necessarily true (Figure 5, following page). Apart from selection bias, case-control studies have a number of other weaknesses. Indeed, several studies of this type produced disappointing results because of various methodological flaws (Cairncross, 1999), as listed below.

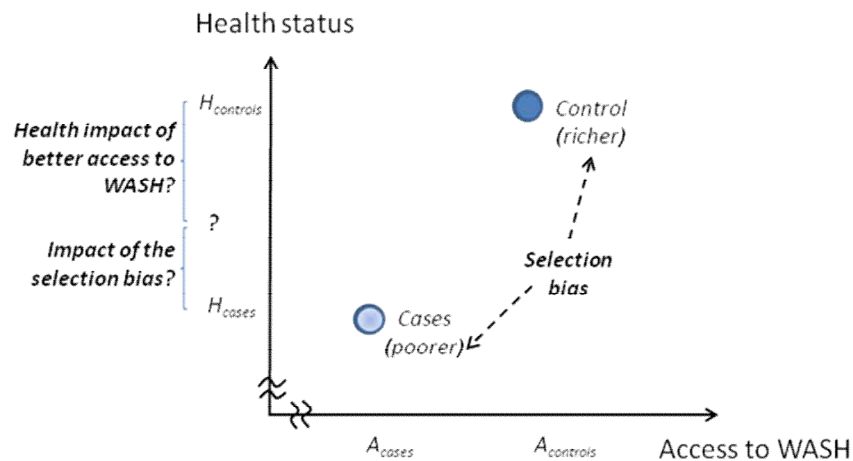


Figure 5. Case-control studies and the selection bias

In 1983, a review of several studies aiming at attributing health benefits to WASH interventions listed 8 common methodological shortcomings (Blum et al., 1983):

- Lack of adequate control: without comparing two groups of people with similar characteristics, evidence-based studies lack rigour (i.e. counterfactual).
- One to one comparison: comparing one village with access to improved WASH infrastructure and one who has not, does not allow evaluators to distinguish whether the measured health differences are due to WASH infrastructure or to other typical aspects of each village (i.e. selection bias).
- Confounding factors: factors having an effect on the outcome simulating a causal relation, which need to be controlled or accounted for in the analysis (e.g. general trend towards better health).
- Health indicator recall: not everybody is able or willing to remember their past health status accurately (making quantitative, statistical methods less accurate).
- Health indicator definition: not clearly defining what differentiates cases and controls (i.e., without a clear definition of diarrhoea, interviewees make their own definitions and report different events).
- Failure to analyse by age: younger and older people often have different behaviours and susceptibilities, which has to be accounted for in health studies.
- Failure to record facility usage: It is the proper use of toilets or safe water sources that has an impact, not just owning one (however, ownership is easy to observe and quantify, while proper use is not).
- Seasonality of health patterns: most water, sanitation and hygiene related diseases are seasonal in large parts of the world - evaluation can produce significantly different results in the dry and rainy seasons.

This study concluded that health impact studies of WASH interventions would always be expensive, difficult to design and conduct and useless unless conducted properly. One of

its recommendations was to use, when possible and feasible, experimental randomised evaluations.

2.4 Experimental evaluations, or Randomised Controlled Trials

To take some of the above comments into account, especially the counterfactual, confounders and selection bias, *experimental evaluations* (also called *Randomised Controlled Trials*, RCTs) take inspiration from medical trials where part of a large group of similar patients is randomly selected to receive an active medical treatment, while the other half is given an inactive *placebo*. In the same way, RCTs randomly allocate the intervention to part of a set of people (*treatment group*), while the rest of the same set (*control group*) does not receive anything. By randomly selecting people for the intervention or the control, the risks of known and more importantly unknown bias is reduced. After the intervention, health indicators can be measured in both groups and compared to deduct the direct impact the intervention had on the treatment one (Figure 6) (Ravallion, 1999 and 2008).

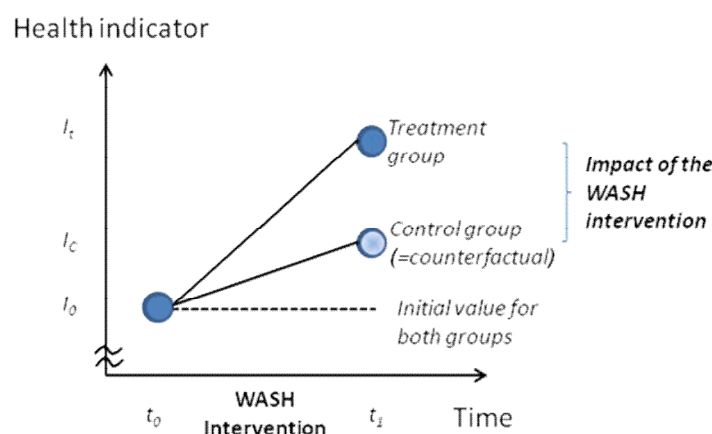


Figure 6. Experimental Impact evaluation or Randomised Controlled Trial (after Ravallion, 2008)

RCTs can be used to evaluate WASH individual or household level interventions. This evaluation method has been used to quantify the health impacts of software interventions such as promotion of handwashing with soap (for instance, Luby et al., 2005), but is more commonly applied to evaluate hardware interventions like point-of-use water treatment (for instance Chiller et al., 2006 or Colford et al. 2005) or urban household water connections (J-PAL, 2007).

RCTs have been advocated as the most rigorous method to evaluate development interventions (Duflo et al., 2003), as randomisation appears the best way to take bias into account and compare beneficiaries to a (similar as possible) counterfactual population. However, RCTs have also been strongly criticised. Even if most suitable to quantify certain impacts, it can be difficult for a single RCT to examine the mechanisms leading to

these impacts and the influence of the context, making it difficult to use the results of such evaluations to generalise findings, or scale up interventions (EES, 2007). A suite of well-designed studies which incorporate a range of data and disaggregated analysis can sometimes provide such information.

If they are theoretically rigorous, RCTs are also practically difficult to implement. For instance, there is no placebo for a latrine or a water connection (Cairncross, 1999), and it can be difficult to properly isolate beneficiary households from non beneficiary ones. They cannot be implemented in country-wide projects and non-standardised interventions. Moreover, allocating WASH interventions' outputs such as water filters or water connections randomly is not always socially or politically appropriate. Interventions usually target specific groups of people (for instance, the poorest) and sometimes, beneficiaries choose whether to participate or not, introducing a de facto selection bias (IEG, 2008).

The best way of doing RCT is to conduct a double blinded, cross-over trial. The first blinding is for the participants, who do not know if they are in the treatment or comparison group (which is given a placebo). The second blinding is for the researcher analysing the data. Cross over consists in, during the intervention, inverting the treatment and the comparison groups (and the treatments and placebos) to make sure their unknown differences does not affect the study (Kirchhoff et al., 1985). No example of such an RCT applied to a WASH intervention could be found as part of this study.

Ethical considerations can complicate RCTs. Medical trials are based on an assumption that researchers do not know which arm of the trial is most effective. Once it is shown that a medicine is clearly effective and has no significant side effects, it is unethical to keep the control group on a placebo, and a data monitoring committee can insist that the trial is stopped early. Development trials are not exactly equivalent, but there are parallel issues. An RCT to test point-of-use arsenic removal kits in a contaminated area could only be conducted ethically if it did not mean withholding a clearly beneficial tool from the control group. It would be ethical to test one method of arsenic removal against another, but what would be the ethics of a trial that relied on the control group having no means of arsenic removal at all? In medical terms, the usual current treatment is known as "standard practice" and most trials are conducted to see if a new treatment is better than standard practice. In the WASH sector, 'standard practice' may be a very low level of service or none at all, but the interests of the control group still need to be taken into account, as the control group is not simply a comparison group but an active part of the trial.

For a combination of these reasons, it has been recognised that RCTs are feasible in only 5% of development interventions (JPAL, quoted in Ba Tall, 2008). A partial solution, at least to the difficulty of allocating the intervention randomly, is to use quasi-experimental evaluation designs.

2.5 Quasi-experimental evaluations

Like RCTs, *quasi-experimental evaluations* also involve comparing a treatment group and a comparison group. The difference is that instead of selecting these two groups randomly in a large set of similar people to avoid biases, quasi-experimental let beneficiaries (or treatment group) be selected or self-select, and try to artificially create an ideal *comparison group* with similar socio-economic characteristics. This can be done through *matching*, i.e. selecting an ideal comparison group using data from a larger survey. For example, data from a national census or a similar survey including socio-economic information on households can be used to select a comparison group which characteristics match the ones from the selected or self-selected treatment group. Alternatively, the comparison group can be selected using a *propensity score* (i.e., the predicted probability of participation given observed characteristics). The closer the propensity score, the better the match, and the less the potential selection bias (Ravallion, 1999).

Because they are often more feasible than RCTs, quasi-experimental studies can be used to evaluate a larger spectrum of interventions, for instance public rural water supply (Jalan et al., 2003) or integrated WASH interventions (Blum et al., 1990). Quasi-experimental evaluations represent a creative way to overcome some of the restrictions of RCTs and they can also be cheaper than RCTs. However, they are still relatively costly, time-consuming and complex. They rely on complex statistical tools, and need relatively large samples and significant amounts of socio-economic data, which might be difficult to collect through surveys (because of budget, staff or time limitations, low population density, long distances, isolated communities). Existing data, such as administrative data from Ministries, or surveys from UN Organisations or NGOs, can be used and adapted through matching and other statistical methods. Bamberger (2006) provided advice on how to conduct quality impact evaluations under budget, time and data constraints. Such methods require highly skilled external evaluators, which also are expensive.

Treatment and comparison groups can only be matched through observable characteristics for which data is available, to avoid any known potential selection bias, while there might also be non-observable or non-measurable differences influencing their choice to participate in the intervention or not (for instance, cultural differences), which would increase the risk of selection bias and reduce the reliability of this method. This can be partially avoided using the *double difference method* (or *difference in difference*). The methods described above are *single difference* estimates, comparing beneficiaries before and after the intervention (time difference only), or comparing beneficiaries with non-beneficiaries (participation difference only, as the two groups are supposed to have similar characteristics). In the double difference method, the treatment and comparison groups (first difference) are compared before and after the intervention (second difference). After the intervention, impact can be quantified by subtracting the difference in the indicators measured between the two groups before the intervention (the selection bias) from the difference in the same indicators measured at the end of the intervention (Ravallion, 2008). This however assumes that both groups are on the same change trajectory, which should

be verified. As an example, this method can be used to quantify the impact of a WASH intervention on the health of a self-selected (hence biased) treatment group, by comparing it to a non-matched comparison group (Figure 7, following page). This method was for instance used by Galiani et al. (2002) to evaluate the health impact of the privatisation of urban water supplies in Argentina, using a quasi-experimental design. The double difference method is also sometimes applied in RCTs.

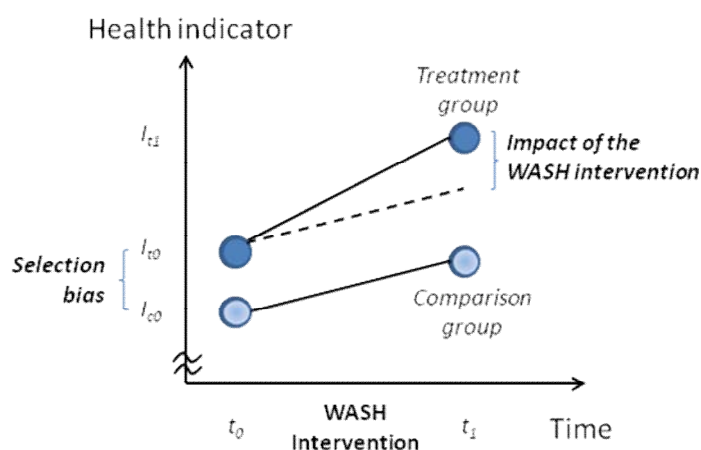


Figure 7. Quasi-experimental evaluation, double difference (after Ravallion, 2008)

In both RCTs and quasi-experimental evaluations, other phenomenon can make it more difficult to compare treatment and the control/comparison groups. Members of comparison groups might be impacted, positively or negatively, by the intervention even if they do not participate directly to it: this is known as the *spillover effect*. For instance, improving sanitation for some households of a community might reduce the contamination of water sources and improve everybody's health, even the comparison group. The comparison group might also be impacted by other interventions, implemented by other agencies, or reproduce the intervention themselves because they like the intervention outputs received by the treatment group: this is called *contamination* or *contagion effect*. In some countries, such as Bangladesh, where numerous agencies implement various projects, it is almost impossible to find large regions where no NGO is already implementing WASH or other interventions. Experimental and quasi-experimental evaluations design might have to include data collection to capture these effects (IEG, 2008). Quasi-experimental studies clearly have value in the WASH sector in some circumstances. However, because of these difficulties, they are not always feasible. The last evidence-based option is to conduct non-experimental evaluations.

2.6 Non-experimental evaluations

An alternative method to experimental and quasi-experimental evaluations are *non-experimental evaluations* (also sometimes called *environmental studies*), which can be used when the study population cannot be chosen randomly nor matched with a comparison group. In such cases, the impacts, or causal relationships of an intervention

can be estimated using statistical methods such as *instrumental variables* (World Bank, 2009). Instrumental variables are variables that matter to participation in the intervention (i.e. the decisions people make to participate or not) but not to its impact. They can be used to establish causality links and quantify impact. For example, if health improvements are observed when an intervention subsidises household water filters, this may suggest a link between household drinking water quality and health: the cheaper the filters get, the more people decide to buy them (i.e. participate to the intervention), and the better their health gets. The price of household water filters might then be used as an instrumental (i.e. intermediate) variable. By studying the *correlation* between fluctuations in price and sales of filters on the one hand (Figure 8a), and the correlation between the sales of filters and health (Figure 8b), the quantitative impact of these filters on people's health can be estimated (Figure 8c). For instance, a study based on non-experimental data was used to evaluate a national programme to control diarrhoeal diseases in the Philippines (Baltazar et al., 2002).

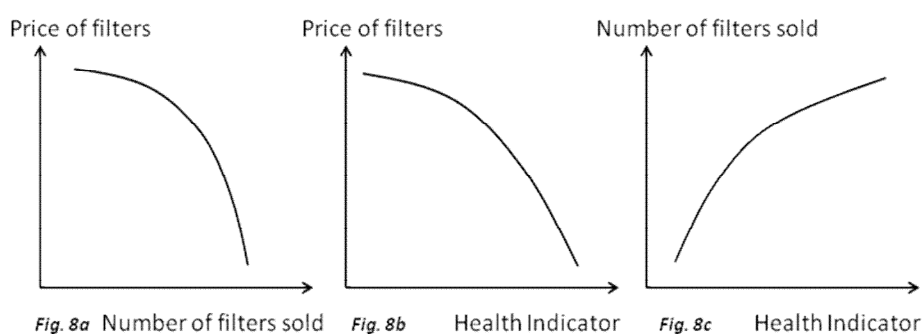


Figure 8. Non-experimental evaluations, instrumental variables and correlations

The various types of evidence-based evaluations described above are currently the dominant focus of development evaluation, including in the WASH sector. They all have strengths and weaknesses, including methodological and practical ones, but their use to quantify the health impacts of WASH interventions, and the use of their findings to orientate WASH development policies have long been, and are still debated. An example is the disagreement between Esrey (1996) and Cairncross and Kolsky (1997, including Esrey's response) on the methodology (biases in a non experimental evaluation using secondary data), interpretation and policy implications (promotion of flush toilets versus latrines) of a multi-country health impact study.

Evidence-based methodologies are well suited for drug trials, in which identical and static treatment and comparison groups can be selected, and all variables can be controlled relatively easily, in order to give meaningful results that can be replicated. Such studies sometimes also take contextual factors into account, but they are not always adapted to development interventions, in which context is much more complex and dynamic. The ideal double blinded, cross over RCTs (see box on page 17) are almost impossible in WASH intervention, as it is difficult to find proper placebos, especially when evaluating new development approaches, and to blind participants and researchers. A major criticism is

that evidence-based evaluations mostly focus on 'what interventions achieved' and not enough on 'how did interventions achieve this', nor 'why interventions are successful or not'. They use experiments to prove or invalidate the predicated impacts, but tell very little about the mechanisms leading to these impacts. They give final results at the end of an intervention, but do not offer many opportunities to 'fine-tune' them during implementation, nor much information on their affordability, acceptability, sustainability and scalability (Clasen et al., 2009). If evaluations are made to replicate or scale up interventions, they should also take *impact heterogeneity* into account, i.e. isolate the specific context of the intervention (geographical, social, institutional or others) that allows the impact, in order to be able to generalise (IEG, 2008). Clearly, these kinds of quantitative evaluations will continue to be used in the development sector. However, the need to learn about on the context and mechanisms of development interventions therefore led evaluators to design what they refer to as *realistic evaluations*.

3. The second paradigm: Black boxes and realistic evaluations

“Our situation is the following. We are standing in front of a closed box which we cannot open, and we try hard to discuss what is inside and what is not.”

Albert Einstein

3.1 The black boxes of development interventions and the theories to open them

To maximise the impact of their interventions, and because of an increasing demand for accountability on their use of resources, development agencies (funding, but also implementing agencies) have focused more and more on impacts and therefore tend to rely on evaluations that use evidence-based methodologies. Such evaluations are nicknamed *black box evaluations*. They give a scientific, quantitative estimation of the impacts caused by development interventions, but little consideration to the intermediate links of the causal chain between inputs and impacts, nor to the process leading from one link to the other. In other words, they put them in a ‘black box’ and give little or no indication as to how and why the intervention led to these impacts. To answer this question, the second paradigm uses theories to open the black box, that is, to look along the causal chain (OECD, 2006).

This first leads to *theory-based evaluations*. These are based on the *programme theory* (also called *theory of change*, or *programme logic*), which documents the causal chain of a specific development intervention, leading to the desired change. The theory is an expression of the intervention logical framework (OECD, 2006), but with a more explicit analysis of the assumptions, and sometimes makes attempts to identify *simultaneous* or *alternative causal strands* (or mechanisms, processes) that could lead to the same change (Rogers, 2008) (Figure 9). Theory-based evaluations analyse the various links of the causal chain and path(s), evaluating whether the theory is or has been realised in practice (OECD, 2006). They are at the same time a process and impact evaluation. By evaluating which assumptions work and which do not, they can suggest corrective action before the end of the intervention. They evaluate positive impacts but can also highlight negative ones, and give a better understanding of how they are generated, and why (Weiss, 2000). The difficulty with theory-based evaluations is however to establish with certitude the causality between the different elements of the theory. For this, they can be combined with evidence-based evaluations, which deal better with the question of attribution.

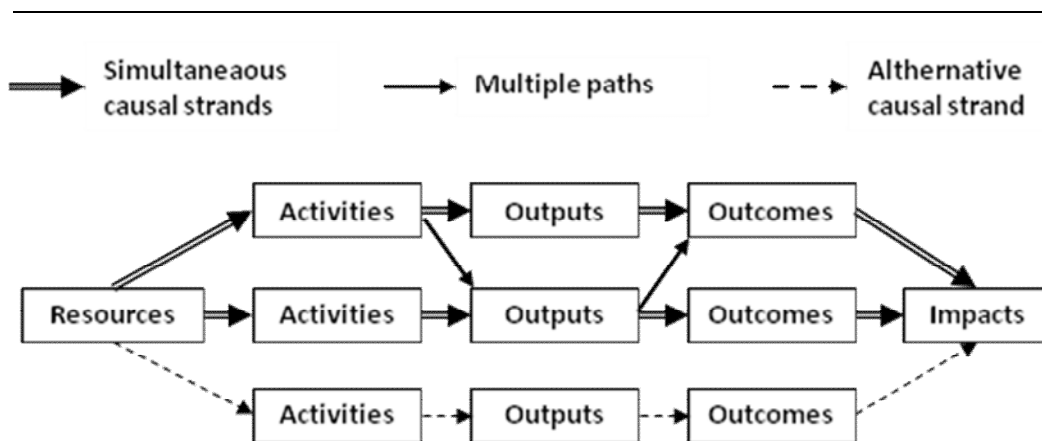


Figure 9. The program theory, in theory-based evaluations, with multiple and alternative causal strands (after Rogers, 2008)

Theory-based evaluations are consistent with the principles of realistic evaluations (NONIE, 2008), but the latter go one step further. Pawson and Tilley, who conceptualised realistic evaluations, believe that theories can explain the mechanisms through which interventions produce changes, but that the linear programme theories (linking inputs, outputs, outcomes and impacts with direct causality arrows) offer a distorted and simplistic view of reality. They believe that the mechanisms through which an intervention leads to its outcomes and impacts largely depend on the context. They call into question the idea of unique and direct causality links between interventions' inputs, outputs, outcomes and impacts, and suggest replacing the causality arrows by *black boxes* to be explored with theories. The black boxes represent an organisation, a social interaction or a social force field in which input is converted into output, output into outcome and outcome into impact, not necessarily through a linear connection (Hospes, 2008) (Figure 10).

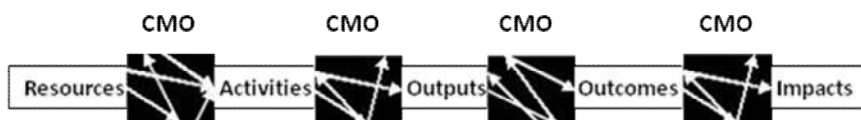


Figure 10. the causal chain in realistic evaluations (after Hospes, 2008)

The most important aspect of a realistic evaluation is therefore the overall contexts in which the intervention takes place. In order to understand the underlying theories of an intervention, evaluators list a series of *CMO configurations* (context-mechanism-outcomes) to explain what happens in, or because of each black boxes, and what possibly determines the outcomes and impacts of an intervention. The *context* refers to the particular implementation environment or characteristics of participants within which a specific mechanism can work. The *mechanism* is the precise way in which an intervention works within a given context to produce a particular outcome. Through logical thinking, experiments or surveys, some CMO configurations can be proved valid and confirmed, or invalid and eliminated. The remaining configurations list the various contexts in which certain mechanisms can produce the desired outcomes. This list can help agencies in deciding where and how they can replicate successful interventions (Tilley, 2000 and Gill et al., undated).

Table 2 highlights the main differences between evidence-based and realistic evaluations.

Table 2. The difference between evidence-based and realistic evaluations (after Hospes, 2007)

	Evidence-based evaluations	Realistic evaluations
Paradigm of development interventions	Interventions stand at the beginning of a results chain	Interventions can be considered as black boxes to be explored
	Inputs → outputs → outcomes → impacts	Context + Mechanism = Outcomes
Type of causality	One-way and single cause and effect relationships	One-way and multiple cause and effect relationships
Objective of evaluations	Measuring effects	Investigating black boxes (i.e. testing theories)
Theory behind interventions and their evaluation	No use of policy theory but use of assumptions and hypotheses of causality (from the logical framework)	Use of policy or programme theories on what happens in black boxes

Realistic interventions have so far mostly been used in European countries in the fields of justice, health and social services and have not yet been taken up by development agencies (Hospes, 2008). Theory-based evaluations are now strongly advocated for use in the development sector (NONIE, 2008 and White, 2006), but no example of their application to the WASH sector was found as part of this study.

However, several attempts have been made in the past to diversify the focus of development evaluations; to try and open the black boxes of WASH interventions by evaluating not only their health impacts but also the contexts and processes leading to them. Two such attempts, the Minimum Evaluation Procedures and the Hygiene Evaluation Procedures, are described below.

3.2 The Minimum Evaluation Procedures

In 1983, in reaction to critics of the health impact evaluations of WASH interventions (including World Bank, 1976 and Blum et al., 1983, op. cit.), the World Health Organisation (WHO), responsible for measuring progress and impacts in the WASH sector during the International Drinking Water Supply and Sanitation Decade (IDWSSD), started to promote the Minimum Evaluation Procedure (MEP).

This evaluation method was based on the idea that health improvements are not a direct result of WASH interventions but result from a long causal chain with various steps (Cairncross, 1999). It acknowledged that water and sanitation infrastructure and hygiene

knowledge do not of themselves create health improvements, but that this comes with their proper management, operation, maintenance, use and practice. The MEP therefore focused not only on the theoretical final link of the causal chain of WASH interventions (impacts), but also on the intermediate ones (outputs and outcomes, i.e. WASH facilities, their use and management; hygiene education and its practice), as described in Figure 11.

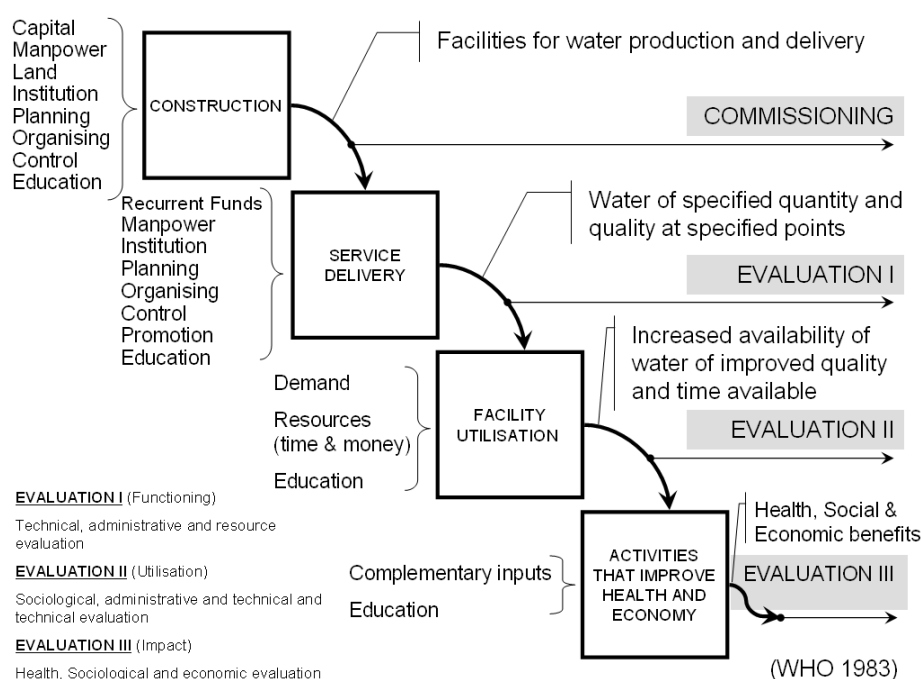


Figure 11. The various steps of WHO's Minimum Evaluation Procedure (Bostoen, 2007, adapted from WHO, 1983)

Moreover, the MEP broadened the focus of WASH evaluations, looking not only for health impacts but also for socio-economic ones (for instance through productive uses of water and time-saving). Another major improvement compared to previous WASH evaluation methodologies was that the MEP relied on simple water, sanitation and health education quantitative but also qualitative indicators to be measured through observations, conversations and small surveys, which made it quick and relatively cheap to implement compared to evidence-based epidemiological studies (Schultzberg et al., 1983).

In spite of its clear relevance, it seems that the MEP was not used after the IDWSSD. The idea behind it was not widely picked up at the time while epidemiological evidence-based evaluations continued to develop. However, a few years later, a new attempt was made to shift from these health impact quantitative evaluations to more qualitative hygiene evaluations.

3.3 The Hygiene Evaluation Procedures

In 1990, a review of the published and unpublished results of the best health impact studies of the IDWSSD concluded that they are not a suitable operational tool for the evaluation or fine tuning of WASH interventions, because of the previously mentioned methodological flaws leading to unpredictable results, which frequently offer no firm interpretation. Instead, this review suggested evaluating the changes in hygiene behaviour that WASH interventions make possible (Cairncross, 1990).

In the spirit of the MEP, and recognising that quantifying the final health impacts of WASH interventions is not enough and that learning about the intermediate links and processes of the causal chain is as important, Almedom et al. (1997) created the Hygiene Evaluation Procedures (HEP). The rationale behind the HEP was that there could not be any significant health improvements from WASH interventions if there were no improvements in hygiene behaviour. Studying changes in hygiene behaviour could therefore serve the same aim as health impact studies, i.e. establishing a link between WASH interventions and health improvements, but would be much cheaper than epidemiological studies, and provide much more information on the intermediate social steps leading to these health impacts. This information could be used not only to evaluate impacts, but also to improve interventions and the processes through which they aimed at changing behaviours. Finally, as the information measured (qualitatively) was closer to the start of the causal chain, it would be easier to attribute the hygiene changes to the project intervention (Cairncross, 1999).

The HEP consisted of a set of qualitative, participatory tools to study hygiene behaviour patterns and their changes throughout WASH interventions. They included tools such as three-pile sorting, pocket charts, semi-structured interviews and focus group discussions to study hygiene behaviours together with beneficiaries, as well as tools such as health walks, structured observations, key-informant interviewing, history lines, community mapping, seasonal calendars and gender role analysis to investigate the context in which hygiene behaviour changes occur (These tools can be found in Almedom et al., 1997). Cairncross and Kochar (1994) wrote various case-studies on the application of these tools. These are still in use today in various WASH interventions, through the HEP or other participatory tools like PHAST (Participatory Hygiene And Sanitation Transformation), two methods showing certain similarities (Almedon, 2003).

3.4 From impacts to outcomes, context and processes, from quantitative to qualitative analysis

The above chronology of the evaluation practice in the WASH sector shows that, in spite of the general interest in the question, there is no real consensus on what should be evaluated in WASH interventions, and how. In broad outline, there are two schools of thought. In the first school, evaluations aim at quantifying impacts (mostly on people's health) and attributing them to WASH interventions through evidence-based,

epidemiological studies, relying on relatively simple, causal chains. Such evaluations are based on sophisticated statistical methodologies, which should be conducted by evaluation experts. In the second school, evaluations focus not only on impacts, but also on outcomes, and the context and processes, relying on more complicated, multiple causal chains. This is achieved through relatively simple quantitative but mostly qualitative methodologies, focusing on health but also on social organisation and changes, and involving beneficiaries.

Quantitative, evidence-based evaluations have clear assets. Their numerical results can easily be stored, analysed, aggregated or averaged, and compared across time and locations (Sijbesma et al., 2008). They enable managers and decision-makers to have summarised indicators of success or failure. But quantitative evaluations also have serious drawbacks. Apart from the methodological flaws discussed above which can produce meaningless results, quantitative indicators may over-synthesise and over-summarise complex information. An example is the experience of the Joint Monitoring Programme (JMP), which was created in 1990 by WHO and UNICEF to monitor national, regional, global, rural and urban water and sanitation coverage. Although very useful to compare objectively countries and regions and track their respective coverage progress over time, the meaning of the quantitative data of the JMP is limited. Universal definitions of concepts like 'having access' to 'improved water sources' and 'improved sanitation' cannot reflect different local realities. Coverage data is available at the national level, but cannot be disaggregated at the sub-national one (Bostoen, 2008).

Some consider WASH qualitative evaluations as less rigorous and scientific than quantitative ones (at least from an experimental point of view). However, they usually are more participatory and give more importance to people's perceptions. This indeed can lead to less objective information, but the numerous qualitative methods available make it possible to check information through *triangulation*. Qualitative approaches might be more appropriate to understand complex human and social processes, within communities but also within the intervention itself. They focus on how these processes work to produce impacts, and how they can be optimised and replicated in other contexts. Because they focus more on the links at the start of the causal chain, qualitative evaluations might not give scientific evidence of the causal links between WASH interventions and their final impacts, but make it easier to attribute the observed outcomes to interventions. . In many cases, they can be coupled with quantitative approaches to give useful, complementary information. In many cases, qualitative data can even be quantified (Sijbesma et al., 2008).

However, both evidence-based evaluations, essentially quantitative, and realistic evaluations, quantitative as well as qualitative, are also criticised. Indeed, they both are based on a systemic approach, and on paradigms of development interventions as more or less sophisticated (simple or complicated) chains of cause and effect relationships, which do not always reflect the complexity of development processes (Mowles et al., 2008, Hospes, 2008).

4. The third paradigm: Complex adaptive systems, complexity evaluations and change assessment

“Everything that can be counted does not necessarily count,
everything that counts cannot necessarily be counted.”

Albert Einstein

4.1 The WASH sector and its increasing complexity

Over recent decades, the WASH sector and its interventions have gradually changed, adapting to new objectives, ideas and research findings. The name of the sector itself, which evolved from ‘water sector’ in the 1970s, to ‘water supply and sanitation sector’ in the 1980s and finally to ‘water, sanitation and hygiene sector’ in the 1990s, reflects this gradual change, in search for more integrated and better, but also more complex approaches.

In the 1970s and 80s, and especially during the IDWSSD, the sector mostly focused on building hardware facilities with the hypothesis that wells, boreholes, latrines, piped networks and sewerage would reduce the time people spent on fetching water, provide them with dignified living conditions and above all improve their health by reducing the incidence of WASH related diseases. However, much of the infrastructure installed during the IDWSSD, by some estimates as much as 40%, broke down after a few years (Schouten, 2006). Although attention to the ‘software’ aspects of water and sanitation systems had enjoyed some attention in previous decades, by the mid- to late- 1990s development agencies came to accept that concrete, pipes and pumps were not sufficient for people to have sustainable and equitable access to safe drinking water, adequate sanitation and improved health. The late 1990s saw significantly increased attention paid within the sector to the social and management components of interventions, especially in the form of establishment and training of water and sanitation committees and associations of users at the community level to operate and maintain infrastructure (Schouten et al., 2007). Community management schemes also resulted from a trend towards decentralisation, through which national governments in many countries gave local communities and district level institutions the responsibility to fund and manage their own water and sanitation infrastructures. They argued that shifting decision making and finances from central to local level would increase a sense of ownership and thereby social participation, better fit local needs and lead to better quality delivery of services (Smits et al., 2005).

This shift in focus away from supply-driven, infrastructure-led solutions towards demand-based social and management approaches with greater emphasis on the role of institutions and economic instruments (Plummer et al., 2007) led to an increasingly complex paradigm of the WASH sector. The WASH interventions themselves became more and more complex, trying to provide both the hardware needed to improve people’s

health and lives, and the software needed for households, communities, local and central institutions to manage and sustain these improvements.

As WASH software issues became more numerous and sophisticated, ranging from hygiene promotion, participation, gender and equity to community management, capacity development, knowledge management, and institutional development, they were gathered under the concept of *water governance*. This is defined as the range of political, social, economic and administrative systems that are in place to develop and manage water resources, and the delivery of water services, at different levels of society (Rogers et al., 2003). With many services still failing to be sustained in spite of increasing efforts to include and improve governance issues into WASH interventions, it is now recognised that the so-called water crisis is essentially a crisis of governance (Plummer et al., 2007).

In parallel with the increasing role of governance in the WASH sector, a new area of enquiry has emerged, *complexity science*. Rather than a formal science, complexity is a collection of ideas and principles (Ramalingam et al., 2008), a new way of seeing the world (Snowden et al., 2007), a body of interdisciplinary knowledge about the structure, behaviour and dynamics of change in complex systems - open evolutionary systems with multiple, strongly interrelated components - including *complex adaptive systems* - complex systems where the components are self-organising and dynamic (after Sanders, 2003).

The weather, rain forests, societies, the global economy, the World Wide Web and the United Nations all fit into the above definition of complex, dynamic, adaptive, constantly changing systems (Sanders, 2003). Some argue the work carried out by development and humanitarian agencies can also be seen as taking place within complex systems (Ramalingam et al., 2008). Table 3 (following page) therefore describes the main characteristics of such systems and attempts to show similarities with the WASH sector, its services and interventions.

Table 3. Links between complex systems and the WASH sector

Characteristics of complex systems (after Snowden et al., 2007)	Similarities with the WASH sector, WASH services and interventions
Complex systems involve a large number of interacting elements.	The WASH sector involves numerous interacting individuals and institutions at different levels, influenced by social, political, legal, technical, environmental, financial and economic factors.
Interactions are non-linear, minor changes can produce disproportionately major consequences.	All the above stakeholders and factors are interrelated with each other. When one changes, it can affect many others.
Complex systems are dynamic, and solutions cannot be imposed; rather, they arise from circumstances. This is frequently referred to as <i>emergence</i> .	Communities, politics, economy and the environment are dynamic systems. People, institutions and WASH services have to adapt each time they change.
The system has a history, and the past is integrated with the present; the elements evolve with one another and with the environment.	Communities, institutions, countries have a history and a culture which have consequences on their ways of working together, and on WASH services.
Complex systems are unpredictable, because the external conditions and system constantly change.	Because of the numerous factors involved and their dynamic nature, WASH interventions and services and their outcomes cannot be predicted.
In a complex system, the agents and the system constrain one another, especially over time, and future cannot be predicted.	Communities, institutions and WASH services constrain each other through governance systems.

This table suggests that, when looked at through the governance lens, WASH services and interventions can also be seen as complex systems. Their numerous interrelated and interconnected stakeholders and influencing factors are dynamic, constantly changing and adapting to each other in an effort to maintain, or regain, equilibrium. In this way initial (often small) effects are magnified through cycles of feedback and reinforcing, or balancing, loops creating virtuous or vicious circles known as *recursive causality* (Rogers, 2008). These *adaptive agents* and their constant change make complex WASH services, interventions and their outcomes and impacts unpredictable; they 'emerge' during the implementation. This is indeed a different way of seeing development, a shift away from the former paradigms of predictable, linear causality (Ramalingam et al., 2008) described above. This new paradigm also has serious implication for their evaluation.

4.2 The WASH sector and the evaluation gaps

In 2004, the Evaluation Gap Working Group, an initiative of the Centre for Global Development, suggested that more resources should be put into more and better shared

studies to evaluate the impacts of development interventions, to know what interventions work under which conditions, what difference they make, and at what cost (Savedoff et al., 2006). Looking more specifically at the WASH sector, it is argued that there are two distinct 'evaluation gaps', an informational one and a methodological one.

A lot has already been published on the impacts produced by WASH interventions but, as mentioned above, most studies focus on health and are of variable quality. Recent *meta-analyses* of past health studies (IEG, 2008 and Fewtrell et al., 2004) suggest that household level water supply interventions (household connections, point-of-use treatment) have the strongest impact on health, while community level water supply interventions (for instance boreholes, wells) produce time savings but only minor health improvements. Improved sanitation also seems to have significant health impacts, but this analysis is based on only few evaluations. There is more evidence that hygiene promotion also improves health significantly. The various health impact evaluations studied in these meta-analyses however show considerably varied results. This heterogeneity of impact is probably related to the design of each intervention and evaluation, and the context in which they were implemented, but little is known about the exact influence of these factors (IEG, 2008).

In spite of these numerous health impact studies, an information gap remains around WASH interventions and evaluations. It is often claimed that improving WASH has impacts on other development issues and can contribute to reduce poverty and hunger, improve education attainment, promote gender and contribute to other Millennium Development Goals (Lenton et al., 2005), but few evaluations have tested these claims (IEG, 2008). Some evaluations do support the theory that there are complementarities between WASH and other sectors suggesting for instance, that maternal education would increase the impacts of WASH interventions. Other evaluations claim health and hygiene education, rather than education in general, significantly increases the impacts of WASH interventions on children's health. Yet other evaluations suggest there is no intra-sectoral complementarity between the different WASH sub-sectors (water, sanitation, hygiene/health education) and argue they are substitutes, with their combined effects being less than would be expected by summing their respective individual impacts (IEG, 2008). These contradictions show there is no consensus on the best way to produce impacts. Another example of the knowledge gap is the lack of documentation around the cost issues of WASH interventions. Very few cost-effectiveness and cost-benefit analysis are available, especially at the local or national levels. Global and regional estimations of the costs and benefits of different water and sanitation intervention scenarios do suggest significant benefits, but these are more due to time savings than health impacts (Hutton et al., 2004).

Filling this information gap could greatly improve development policies. Knowing better whether water, sanitation and hygiene interventions have greater impacts alone or combined, whether they are complementary to or substitutes for other types of interventions could guide WASH development investments. Cost-effectiveness and cost-

benefit analyses would also enable decision makers to choose between different types of interventions and optimise the impacts of their investments (IEG, 2008).

But even more important to the WASH sector is the methodological evaluation gap. The evaluation methods currently used in the sector – mostly evidence-based – can quantify impacts to prove what works and what does not. But these methods can mostly be applied to evaluate the impacts WASH interventions have at the beneficiary level, and are not appropriate for all types of interventions.

RCTs can prove health or socio-economic impacts of household level WASH interventions (point of use treatments, household connections, handwashing campaigns or improved sanitation) but, as explained above, are rarely feasible. Quasi-experimental evaluations offer more possibilities, but again, are essentially useful to quantify impacts at household level, while WASH interventions often work at the community level (boreholes, wells, community hygiene campaigns, community coverage of improved sanitation). Taking large enough treatment and comparison samples of communities rather than households would be too costly and too difficult to implement. WASH evidence-based evaluations can be compromised by the spillover and contamination effects, and cannot explain impact heterogeneity.

Moreover, it is not enough to know whether an intervention works or not. Information is also needed on where and when it works, why and how it works, and how to scale it up and sustain it, on which there is currently little information. The theory-based or realistic evaluation approaches might give more information on which type of intervention can work in which context, but such evaluations, studying both impacts and process, have not yet been applied to the WASH sector (IEG, 2008).

Finally, both evidence-based and theory-based methods rely on systemic paradigms of WASH development interventions. They conceive interventions as single or multiple chains of steps and factors linked together by causes and effects. They can be adapted to certain WASH interventions, but not all. If WASH governance and the interventions trying to improve it are considered as complex systems, in which all factors and stakeholders are interrelated, continually constraining each other and adapting to each other, these systemic, causal paradigms are not fully valid anymore. Some attempts have been made to use theory-based methods to evaluate complex interventions. In some cases, some aspects of a complex intervention can be understood as simple or complicated, and therefore can be dealt with a programme theory. In other cases, complex interventions can be modelled with a flexible or 'emergent' programme theory, i.e. a theory that evolves as the intervention develops, and as new findings and outcomes emerge. It is however recognised that programme theories, however detailed and flexible, cannot generate performance measures that can be used formulaically to modify implementation and improve performance of complex interventions. For complex interventions, thoughtful indicators have to be used thoughtfully (Rogers, 2008).

It is recognised that there is a lack of knowledge on the impacts of interventions such as software reforms and community-based management (IEG, 2008). New methodologies might therefore be needed to evaluate such interventions.

4.3 The third paradigm: complex systems, complexity evaluations

In the complexity paradigm, interventions and their impacts depend on numerous interrelated factors and stakeholders, which are constantly changing. Complex interventions therefore cannot predict their impacts, and cannot produce them systematically. They can control their resources and the resulting activities and outputs, but can only facilitate the changes, or outcomes, needed for the intervention partners to produce impacts. The links between these different components of complex interventions can themselves be understood as complex, as illustrated below by black boxes (Figure 12). Resources, activities and outputs might therefore need to be adaptive and responsive.

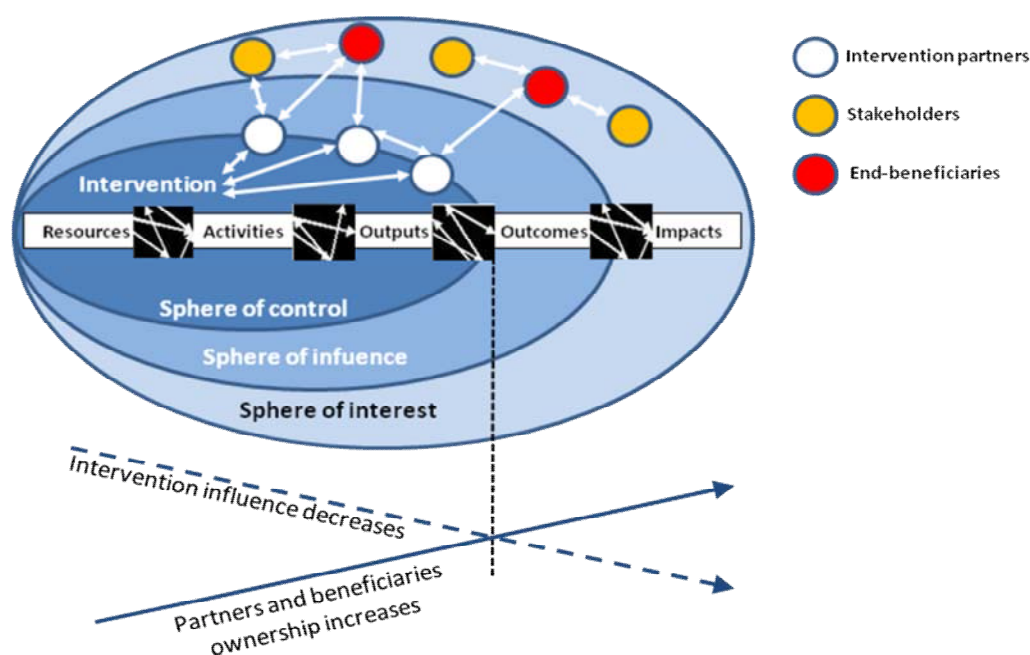


Figure 12. The spheres of control, influence and interest of complex development interventions (after Hearn, 2008)

WASH interventions including governance issues fit particularly well in the above illustration of the paradigm of complexity interventions. If the construction of WASH infrastructure such as wells, boreholes and latrines (i.e. activities and outputs) might be under the control of funding and implementing agencies, their long-term, equitable and sustainable management is not. This responsibility falls on their local partners, e.g. WASH committees at the community level, as well as local governments and Ministries at district and central levels. As project interventions do not control these institutions, they can only provide them with tools, for instance through capacity-building and institutional

development activities, to influence the changes needed for them to manage WASH systems and services in an equitable and sustainable way.

Rather than focusing on impacts that cannot be predicted, *complexity evaluations* focus on outcomes, i.e. changes, as well as the complexity of the links between the various components of an intervention. They try to foresee (when possible) the changes needed to be achieved to produce impacts, and use various methodologies to observe, document and track them, as well as the complex processes leading to them. Some methodologies can also be used to spot and analyse emergent changes that could not be predicted. For this, complexity evaluations mostly rely on qualitative methods, some of which will be presented in Section 5 of this TOP.

This approach is quite different from the evidence-based and realistic evaluations described above. Table 4 summarises and compares the main characteristics of these three types of interventions.

Table 4. The differences between evidence based, realistic and complexity evaluations (after Hospes, 2007)

	Evidence-based evaluations	Realistic evaluations	Complexity Evaluations
Paradigm of development interventions	Interventions stand at the beginning of a results chain	Interventions can be considered as black boxes to be explored	Interventions are adaptive systems
	Inputs → outputs → outcomes → impacts	Context + Mechanism = Outcomes	Outcomes are emerging and quite unpredictable
Type of causality	One-way and single cause and effect relationships	One-way and multiple cause and effect relationships	Two-way and multiple cause and effect relationships
Objective of evaluations	Measuring effects	Investigating black boxes (i.e. testing theories)	Exploring complexity
Theory behind interventions and their evaluation	No use of policy theory but use of assumptions and hypotheses of causality (from the logical framework)	Use of policy or programme theories on what happens in black boxes	Starting point is that policymaking is dynamic and interactive

This TOP has shown that there are several ways of conceptualising development interventions, and consequently numerous tools for evaluating them. It has also highlighted that different experts and organisations have different views on what are the best evaluation tools, which can be used when and so on. The purpose of this TOP is not to

'settle' these arguments but to demonstrate areas of debate and show how different types of evaluation can be more appropriate for different situations and for answering different questions. However, other important questions also need to be addressed.

4.4 Why and for whom should we measure the WASH sector's successes and failures?

So far, the main driver for measuring the WASH sector's successes and failures has been the need for funding and implementing agencies to be transparent and accountable, and to show evidence that their WASH interventions are producing impacts, essentially on the health and sometimes livelihoods of their end-beneficiaries. Indeed, private people, taxpayers or institutions who financially contribute to WASH interventions want to get 'value for money', hence the question 'how much have you have achieved?' It is only fair that agencies should be accountable to those who fund them and show results. However, the current dominant evaluation methodologies are not always adapted to answer this question, especially for complex WASH interventions including governance issues, and evaluating projects only from the standpoint of accountability to the funders is a missed opportunity to learn more about the complexity of the sector and its governance. Moreover, with improved governance increasingly seen as the key to equitable, sustainable and WASH services that address the needs of whole populations, this learning is critically needed to achieve development targets such as the Millennium Development Goals on water, sanitation and health.

Evaluations should therefore not only try to 'prove' impacts, but also look at how to 'improve' interventions. 'Proving' is about demonstrating that change is actually taking place: 'improving' refers to a continuous process of learning about how change occurs, how it can be supported and sustained (after NEF, 2009). The proving part of evaluations is mostly for the benefit of funding and implementing agencies. The improving part could enable all development stakeholders, including funding and implementing agencies but also local partners, institutions, beneficiaries and the public, to better understand each other, and learn how to work more effectively and more efficiently towards common goals and better services.

These various stakeholders however have different interests, different expectations, and different understanding of evaluation results. According to Habicht et al. (1999), the main objective of an evaluation is to influence decisions, and the complexity and precision of an evaluation should depend on who the decision makers are and on what type of decisions they will take as a consequence of the findings. If evaluations are not only for funding and implementing agencies but also aim to benefit local partners, institutions, beneficiaries and the general public, evaluators will have to adapt the type and level of information they present, as well as the language with which it is presented.

Current evaluation practice in the development sector in general, and in the WASH sector in particular, is not adapted to such diverse and ambitious objectives. New approaches are

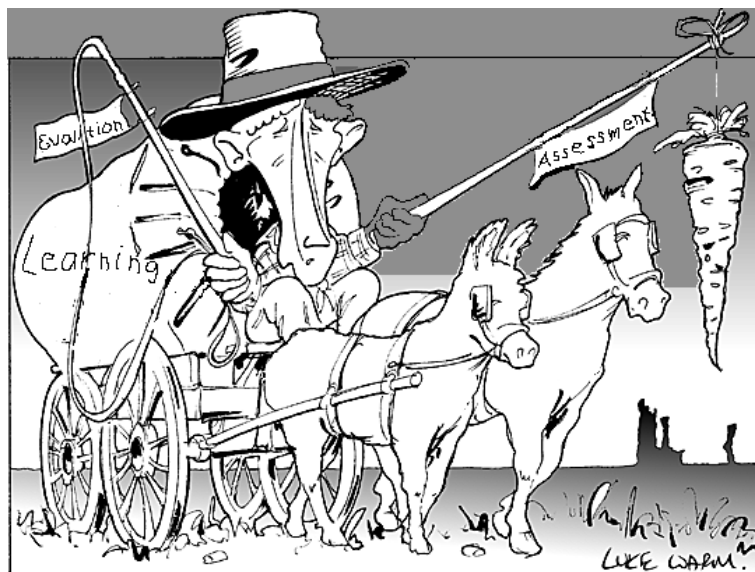
therefore needed to give a broader picture of the changes achieved by complex WASH interventions, as well as the processes through which they achieve them, to learn, influence but also empower many different decision makers, from the individual and community levels to the district, national and international levels.

4.5 Evaluation or assessment?

As with most publications on this subject, this TOP has so far only discussed evaluation. A minority of other documents and organisations however prefer the term assessment, which has a related but different meaning. It is interesting to note that the most different and clearest definitions of these terms which were found as part of this study come from the Education sector, to differentiate *assessment* and *evaluation* within learning processes. Table 5 and Picture 1 (following page) summarise and illustrate these definitions which, although not universal, will be used in the rest of this document.

Table 5. Comparison between assessment and evaluation (after Parker et al., 2001 and Straight, 2002)

	Assessment	Evaluation
Timing and purpose	Ongoing, to improve learning and future performance	Final, to judge the merit or worth of a performance against a pre-defined standard
Focus	Process-orientated, how learning is going	Product-orientated, what has been learnt
Setting Criteria	Both the person being assessed and the assessor choose the criteria.	The evaluator determines the criteria
Relation	Learn from each other	Trying to find fault with the intervention, or vice versa, with the evaluation; to "beat each other out"
Control	The person being assessed, who can choose to make use of assessment feedback	The evaluator, who can make a judgement on the person being assessed
Depth of Analysis	Thorough analysis by answering questions such as why and how to improve future performance	Calibration against a standard
Response	Positive outlook of implementing and action planning based on (joint) learning	Closure with failure or success



Picture 1. Differentiating assessment from evaluation as continuous improvement tools (in Parker et al., 2001)

This suggests that the difference between evaluation and assessment is similar to the one between proving and improving. It is therefore not surprising that evaluation has so far been the dominant term. But this also highlights how evaluation and assessment have different objectives and need different methodologies. As evaluation is to make a judgement, and give a standard value to a performance to compare it with others, it is not surprising that it uses more top-down, quantitative (although not exclusively) approaches focused on impacts. An assessment is more oriented towards improvement on a case by case basis, and uses more participatory approaches, quantitative but also qualitative, focusing on changes and processes of changes. This is not about “bad cop” (evaluator), “good cop” (assessor). Clearly both are interested in the impacts of interventions, if these are evident. However, evaluators are primarily focused on evidence of outcomes, while assessors are more interested in processes and how they can be improved, as these can ultimately improve outcomes.

Dominantly quantitative evaluation and more qualitative assessment both have their proponents and critics. It is argued here that they both are useful for the purpose of learning and improving performance, including in the development and WASH sectors. They serve different, but complementary objectives. They should therefore be used together, to maximise what can be learnt from the successes, as well as the failures of WASH interventions.

The above sections presented various evaluation methodologies. The following sections concentrate on change assessment methodologies that can be used in complex WASH interventions, to better understand the complex relationships between the various partners and components of an intervention, and how and why outcomes emerge or not.

5. Some examples of participatory tools to assess change in complex WASH interventions

“The only constant is change.
Change is the only constant.
Change alone is unchanging.”

Heraclitus

Neither participatory tools nor change assessment are new in the WASH sector. Various participatory methodologies have been widely used in the past, for instance the Participatory Rural Appraisal (PRA) (World Bank, 1996) and its numerous adaptations, PHAST (Participatory Hygiene And Sanitation Transformation) (WHO, 1998), or the HEP (Hygiene Evaluation Procedures) (Almedon et al., 1997). References also already exist on participatory impact assessment methods. Catley et al., (2008) describes tools for livelihoods projects that can be adapted to WASH interventions while Narayan, (1993) details participatory evaluation and change assessment, specific to the WASH sector.

This section does not present a complete list of all the existing participatory assessment tools. Those mentioned above are already well known to most development practitioners and well documented, and the references at the back of this TOP can be used for more information. Rather, it will focus on more recent or less widely used participatory assessment methods, mostly qualitative but also including some quantitative components, that can be used to assess changes in complex WASH interventions. Each method is briefly presented, including a presentation of the methodology and a short case-study describing how it has been used in WASH governance interventions.

Other innovative assessment tools exist that have not yet been used in the WASH sector. For instance, several qualitative assessment methods have been developed, for instance by Rick Davies (Evolving storylines, Hierarchical Card Sorting, Basic Necessities Survey, Weighted checklists, see Davies, 2009), or by Boru Douthwaite (Impact Pathway Analysis, see Douthwaite et al, 2008). Those which have already been applied in WASH governance projects are described below while others that might also prove useful to the WASH sector can be checked in the above sources (see References and TOP websites). IRC (International Water and Sanitation Centre) is interested in receiving feedback on innovative assessment tools and methodologies, and how they were used in WASH interventions (see contact details on the last page of this TOP).

5.1 Balanced Score Card

The *Balanced Scorecard* (BSC) technique is a quantitative and qualitative assessment method originally designed for the private sector, later adapted to assess public services (Estis, 1998). It involves identifying and measuring a few key objectives in four

perspectives of performance (see figure below) that counterbalance one another. The BSC therefore provides a broader picture of a public agency's performance and sustainability than its technical records or financial accounts alone. It encourages managers to optimise the performance of their agency by balancing the four perspectives over the longer term. Importantly, this method focuses on ongoing improvement rather than impact. The *operational efficiency perspective* emphasises the need for the agency to deliver its

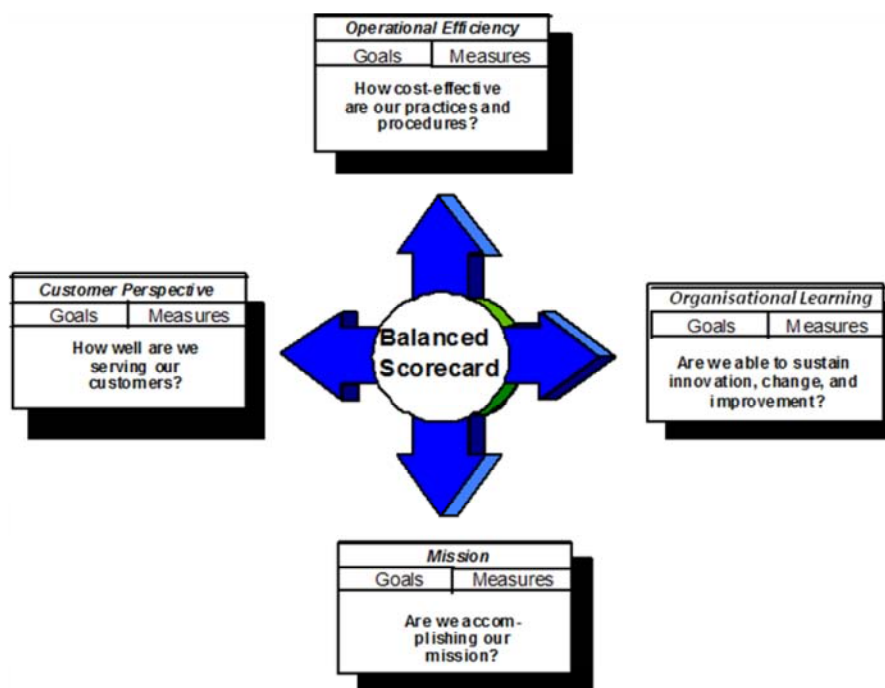


Figure 13. The Balanced Scorecard, adapted to the public sector (Estis, 1998).

services at the lowest possible cost. This can be measured with unit costs, changes in expenditures over time, or administrative to direct service costs ratios. This perspective is a counterbalance to the other three, making sure that the agency is not spending too much to accomplish its goals.

The *customer perspective* focuses on the stakeholders the agency serves and how satisfied they are with the service, its timeliness, quality and charged cost. These can be measured for instance with waiting or response times, downtimes before equipment is repaired and overall customer satisfaction. The customer perspective is a counterbalance to the operational efficiency perspective, as serving customers properly costs money.

The *mission perspective* focuses on the primary public service mission that a government agency has been created to accomplish. The way it can be measured depends on the nature of the mission. This perspective is a counterbalance to the operational efficiency and customer perspectives. It reminds managers that the agency should not pursue efficiency and customer satisfaction at the expense of the other parts of the larger mission.

The *organisational learning and improvement perspective* focuses on the capital, labour and processes investments an agency needs to make to prepare its future and ensure its sustainability. Examples of indicators could include the level of investment in staff training, the number of new innovative procedures or technologies developed and tested. This perspective is also a counterbalance to the operational efficiency perspective (Estis, 1998).

Using the Balanced Scorecard to assess WASH institutions in Uganda		
WaterAid proposed to apply the BSC to assess Ugandan district water agencies (Kanyesigye et al., 2004, Slaymaker et al., 2004), as shown in the table below.		
Table 6. A district Balanced Scorecard for the Ugandan Water and Sanitation sector (Kanyesigye et al. 2004, Slaymaker et al., 2004)		
Performance area	Goal	Possible performance measures in WSS
Achievement of mission Finding the extent to which objectives and goals are being realised	<i>Sustainable safe water supply and sanitation facilities, based on management responsibility and user ownership, within easy reach of the rural population by the year 2005 and with an 80%-90% effective facility use and functionality rate. Eventually increasing to 100% of the rural population by the year 2015.</i>	<ul style="list-style-type: none"> • District water coverage • District Water Point Density • Average and relative variation in WPD • Functionality of existing water points • Household Latrine C coverage • Institutional Latrine Coverage
Operational efficiency The value for money of services being provided	<i>Water and sanitation services delivered efficiently to the population, using appropriate low cost technologies where possible</i>	<ul style="list-style-type: none"> • Unit costs of constructing different facilities • Average per capita investment cost • The technology mix (proportion of low cost technologies) • Collection of capital contributions
Customer Perspective How well are customers being served?	<i>The population is actively engaged in decision-making over WSS facilities, managing and using high quality sustainable water and sanitation facilities</i>	<ul style="list-style-type: none"> • Community engagement in the planning process • Results from the sustainability snapshot • Results from the sanitation snapshot • Water quality & quantity • Service
Service and improvement How has and what is the likelihood that services will improve?	<i>Local government is making improvements in the delivery of efficient, equitable and sustainable water and sanitation services</i>	<ul style="list-style-type: none"> • Improvements in safe water coverage over the last two years • Improvements in equity over the last two years • Improvements in unit costs over the last two years • Quality of workplans
WaterAid suggests that the BSC could be used by district water offices to assess the WASH performance in each sub-county, as well as by the Directorate of Water Development to assess the district water offices. This bottom-up assessment would help to identify good and bad practice in local governments, and to allocate more resources to sub-counties and districts performing poorly. Results could be published to inform people on the performance of their local authorities, and as an incentive for managers to improve.		

5.2 Most Significant Change

The *Most Significant Change* (MSC) technique is a form of participatory monitoring and assessment, in which the various project stakeholders are involved in deciding what changes they want to have recorded and in analysing the data. MSC is a simple, qualitative method which does not require pre-defined indicators, either quantitative or qualitative.

'Significant change stories' are collected on a regular basis from samples of intervention stakeholders (beneficiaries, partners, staff, etc). A group of designated staff and stakeholders gathers to read the stories aloud, have in-depth discussions about the value of the reported changes, and select the most significant one. To have good learning value, this exercise must be repeated regularly, and possibly at different levels of the same intervention (community, district, province, etc.). Choosing the 'most significant' change is clearly a critical factor that will influence how the intervention is seen.



Picture 2. The Most Significant Change method is based on storytelling and selecting (Davies et al., 2005)

The MSC technique can be implemented in ten steps:

1. Getting started: appointing a responsible team of staff and stakeholders, and getting them familiar with the approach
2. Establishing 'domains of change' on which all stories must focus
3. Defining the reporting period, with deadlines for collecting and reviewing the stories, as well as date by which the most significant one is sent to the upper level if necessary (e.g. from community level to district level)
4. Collecting stories of change
5. Reviewing the stories with the appointed team
6. Providing stakeholders with regular feedback about the review process
7. Setting in place a process to verify the stories if necessary
8. Quantification of the changes, when possible and if necessary
9. Conducting secondary analysis of the stories en masse, to find similarities, trends, or other patterns
10. Revising the MSC process

Unlike traditional techniques that focus largely on monitoring activities and outputs, MSC focuses on monitoring intermediate outcomes and impacts. This technique can be implemented throughout the project cycle and provide information to help people manage and improve the programme, identifying unexpected changes. It also contributes to evaluation because it provides complementary data on diverse impacts and outcomes that can be used to help assess the performance of complex interventions.

MSC is simple to understand, and adaptable to any culture and to any level of an organisation or intervention: everyone can tell stories. It encourages analysis as well as data collection, contributing to building a rich picture of what is happening, rather than an overly simplified picture where organisational, social and economic developments are reduced to indicators and numbers. Despite its apparent simplicity, MSC can also be challenging to implement properly.

MSC is also better suited to certain types of interventions than to others. Conventional, quantitative monitoring and evaluation may be sufficient and less time consuming in interventions in which outputs, outcomes and impacts can easily be quantified (for instance, distribution of point-of-use water treatment technologies), where changes are expected and for retrospective evaluation. In more complex interventions, with numerous organisational layers and emergent outcomes, focusing on social change and using participatory methods, MSC can add learning and complexity dimensions to on-going conventional monitoring and evaluation for accountability (Davies et al., 2005).

Most Significant Changes in 3 projects implemented by ADRA in Laos

ADRA Laos (Adventist Development and Relief Agency) used the MSC technique on a pilot basis in 3 WASH and health projects, and commissioned an evaluation of its use, challenges, opportunities and lessons learnt. Positive and negative MSC stories were periodically collected and analysed at several levels of the projects (monthly at the field, regional, and national levels in Laos, and bi-annually at the international level by the donor in Australia), using pre-defined formats. Positive MSC stories included “reduced time and labour for women”, “growing more vegetable”, “boiling water, using garbage and mosquito nets” or “reduced incidence of dysentery”, while negative ones included “queuing time”, “bad water taste” and “inadequate drainage at water points”. In spite of apparent simplicity of the MSC technique, ADRA found it challenging to collect stories objectively, analyse them in a collective and participatory manner, provide proper feedback at the various project levels and make adequate use the findings to improve projects as well as for individual and organisational learning. The process however proved very useful, with project staff and beneficiaries focussing more on change. One of the most useful aspects of using the MSC technique was discussing and analysing the stories (with project staff and stakeholders, but it was recommended to include beneficiaries as well), as this process fostered an organisational culture of learning. The evaluation of this pilot project recommended that ADRA and other agencies use the MSC technique to serve their learning and development work (Willetts, 2004 and Willetts et al., 2007).

5.3 Social Network Analysis

Development interventions usually involve a variety of stakeholders, interacting between each other and the larger environment. These can be seen as networks of people (individuals, communities, organisations, institutions, companies) and things (interventions, activities, documents, events), connected by relationships (transmission or exchange of information, money, goods, influence, friendship, etc.). *Social Network Analysis* (SNA) is a body of social science methods that can be used to model and analyse the structure of these networks, as well as the nature and frequency of relationships between members. SNA can be used in a variety of development interventions, but is more useful in large interventions with multiple objectives, many stakeholders and complex relationships.

A network can be modelled through a *matrix* or a *diagram*. Network diagrams can describe complex structures of relationships between actors or networks to plan improvements or, by repeating the analysis, to assess changes. Mathematical measures can also describe the overall structure of a network or the position of a member within a network.

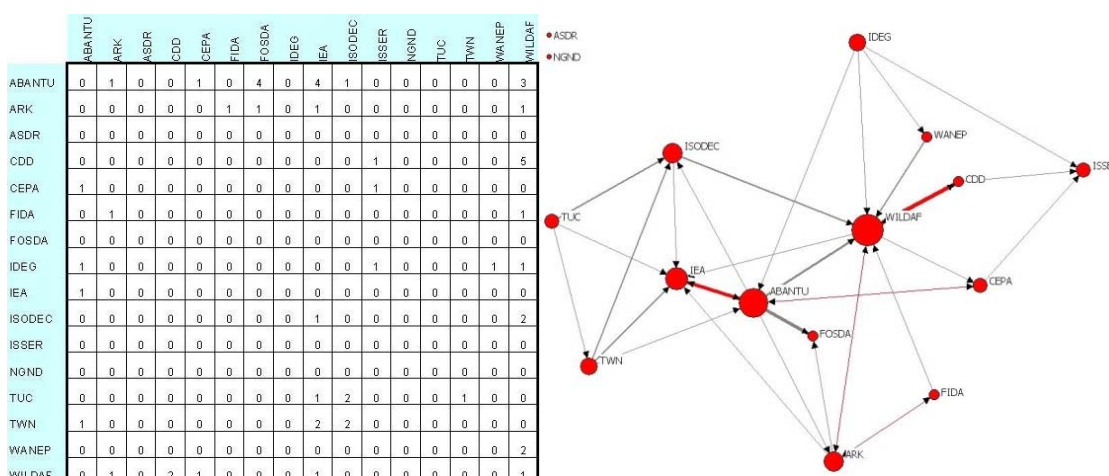


Figure 14: Matrix and diagram of the same network of NGOs in Ghana (Davies, 2008).

Figure 14 models a network of 16 NGOs in Ghana. In the matrix, each cell marks the number of times a specific relationship was mentioned in an NGO report to their donor. The diagram derived from this matrix illustrates some characteristics of this network – a cluster of NGOs linked together (left), two isolated NGOs (top left), NGOs with many connections (centre). Other aspects of NGO relationships could have been used, giving a different picture. The choice of what goes in a matrix or a diagram is therefore important.

These tools can be used to describe the nature and structure of relationships between members, including the position of members within a network (centrality, closeness to other members, ‘betweenness’), the positions of several stakeholders (reciprocity of relationships, structural equivalence with the rest of a network), and features of the whole network (size, isolated components, clusters, density, cohesion). Inquiries can also explore incoming links that provide resources or influences on the group, or outgoing links whereby

the group is able to influence others. Network diagrams can also be used to analyse inter-relationships within a portfolio (e.g., of NGOs funded by a donor), or to highlight possible pathways of influence (e.g., how NGOs might influence government policy). It can hence be used to develop and assess network strategies. As with other forms of social research, it is important to seek the views of actors and to have a good understanding of the context.

SNA tools can be used to clarify Logical Framework descriptions of development projects with multiple outputs, outcomes and associated indicators. The complexity of expected relationships between the various project components can be captured in SNA matrices, developed in a participatory manner e.g. via stakeholder workshops.

A social network approach to thinking about development interventions emphasises the fact that most social change takes place through human relationships. It is not only the character and capacity of individuals (and organisations) that matters, but the structure of relationships between them (Davies, 2008).

Social Network Analysis in the Katine Community Partnerships Programme (KCPP) in Uganda (Rick Davies, MandE)

Katine, in Soroti county, has some of the worst health and living conditions in rural Uganda. The aim of the Katine Community Partnerships Programme (KCPP) is to improve access to water, healthcare and education, to improve livelihoods and to empower men and women to take charge of their lives through better governance. The focus is on capacity development and training (e.g. on basic rights to health, water and education and on planning and budgeting) and on gathering and using local data to influence resource allocation.

Social Network Analysis was used in KCPP to map the relationships between more than 50 stakeholder groups. By importing AMREF (African Medical and Research Foundation) database on these various stakeholders and their numerous interrelations into social network analysis software, it was possible to generate a single network diagram.

The map is being refined during periodic monitoring visits. The external evaluator is asking group members what other groups they belong to, and which of those relationships are most relevant. Inquiries about incoming (resource) linkages highlighted the importance of links with groups not yet included, notably various churches. Other questions could be asked: about overlaps in membership of groups (especially office holders), which groups are peripheral or central, and about clusters of stakeholders.

The project design initially emphasised the development of new ways of doing things: their adoption by others in the community was not articulated in detail. A simplified network model has highlighted at least four pathways whereby government partners could be influenced to make wider use of innovative practices. The next monitoring visit will ask how effectively each of these pathways is being exploited. Subsequent visits will examine if the simplified social network map has made any difference to how AMREF promotes its models of practice in the areas of health, education and agriculture.

5.4 Outcome mapping

Outcome mapping is a process-oriented learning and improvement method focusing on the changes an intervention can facilitate and influence. It defines outcomes as the changes in the behaviour, relationships, activities, or actions of the people, groups, and organisations with whom an intervention works directly. Outcome mapping encourages implementers and partners to reflect on the outcomes or changes they want to achieve through an intervention, and provides them with tools to define, monitor and assess these changes while implementing an intervention, as well as to evaluate them once activities are completed. This tool is particularly well adapted to understand the processes of change (how and why things change) underlying complex interventions involving multiple stakeholders at different levels.

As described in Figure 15, outcome mapping is composed of three stages and twelve steps. For each stage, the Outcome Mapping Manual (Earl et al., 2001a), provides tools and worksheets to assist interventions to organise and collect information on their contributions to desired outcomes.

Outcome mapping is usually initiated through a participatory workshop with all the intervention *boundary partners* (the individuals, groups, and organisations with whom the intervention interacts directly to effect change and with whom the intervention can anticipate some opportunities for influence). The first stage, *intentional design*, enables the partners to reach a consensus on the macro-level changes the intervention should support, and to plan the strategies it will use to influence these changes and monitor the

progress towards achieving them through *progress markers* (a set of graduated indicators of behavioural changes for a boundary partner that focus on depth or quality of change). This first stage helps answer four questions: why does the intervention want to influence changes (step 1)? How will it do it (steps 2, 6 and 7)? With whom (step 3)? And what changes will be monitored and assessed (steps 4 and 5)?

The second stage, *outcome and performance monitoring*, uses a set a qualitative tools to monitor and assess changes in the boundary partners using the progress markers. These



Figure 15. The three stages and twelve steps of outcome mapping (Earl et al., 2001).

tools clarify directions with boundary partners and monitor outcomes (step 9: *outcome journal*), strategies and activities (step 10: *strategy journal*) and organisational practices (step 11: *performance journal*). They enable the intervention to reflect on and improve its performance in the achievement of change, and to collect data on the results of its work with its boundary partners.

The third stage, *evaluation planning*, uses the results of outcome mapping activities to set evaluation priorities in an evaluation plan (Earl et al., 2001a and 2001b).

Outcome mapping can also be integrated with the logical framework approach (Roduner et al., 2008).

Outcome mapping in Ethiopia in the RiPPLE project (Ewen Le Borgne, IRC)

RiPPLE (Research-inspired Policy and Practice Learning in Ethiopia and the Nile Region) aims at promoting uptake of research findings in policy and practice through a Learning and Practice Alliance (LPA). With six district (woreda) LPA platforms, three regional ones and a national Forum for Learning on Water and Sanitation (FLoWS), the project relies heavily on influencing sector stakeholders to carry out joint research, discuss and analyse key issues and take up research results and recommendations. With respect to its focus on behaviour change, the outcome mapping (OM) methodology appeared to be an appropriate approach to measure the degree of interest, communication and cooperation with other sector stakeholders in Ethiopia and the Nile region, as well as the uptake of research results and the extent to which LPA members would promote the LPA and research. Ten platform coordinators were trained in the OM methodology, which was adapted to fit RiPPLE's needs: outcome journals, strategy maps and strategy journals were used, but monitoring priorities and organisational practices were left aside. Progress markers included for instance "Regularly attending LPA meetings", "Collaborating with other WASH actors", or "advocating that generated information will be used". RiPPLE coordinators have been filling the outcome and strategy journals and maps for two to three different boundary partners (BP) in each region: LPA members, active NGOs in the region and specific governmental bureaus. The national level coordinator has been monitoring the Ministry of Water Resources, WASH forum and donor advisory group.

Despite some initial confusion in using the methodology, RiPPLE coordinators have been supported by the monitoring team and attended an LPA facilitation training course that included monitoring and the use of OM. They have become very clear about their facilitation role in the change process in the LPA, and have been able to devise a better plan to achieve a better coordination in the LPA and improved research results that are used by governmental agencies and local implementers. They have all taken up specific issues of integrating LPA platforms, focusing on policy engagement and discussing research results with policy makers in face-to-face discussions. RiPPLE is running until 2011, and outcome mapping, in conjunction with other tools (such as interviews and social network analysis), is proving a very useful approach to promote social change and increased cooperation in the WASH sector in Ethiopia.

5.5 Process documentation

The concept of *process documentation* has evolved over recent years as a relatively informal but dynamic way of monitoring the processes within an intervention and the changes that result from it. Initially called process monitoring (Zimmermann et al., 1996), it uses a flexible set of assessment tools to track, document and analyse meaningful events related not only to an intervention but also to its larger context (socio-cultural, geographical, historical, political, institutional, etc).

Development interventions are often designed as straight lines towards the changes they want to achieve. But, as illustrated in the figure below, many events force a project to adapt its course when it is implemented in a dynamic environment.

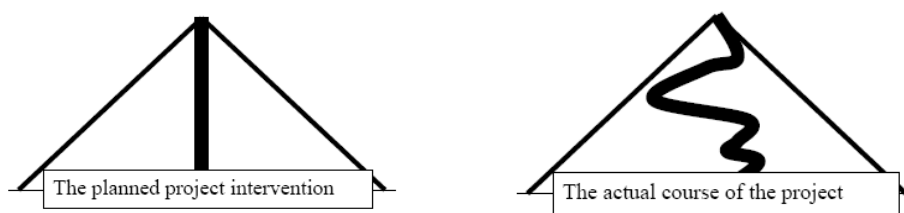


Figure 16. How a straight intervention adapts to a dynamic environment (Schouten, 2007).

Process documentation is a flexible approach to observing and learning from the real world of complex projects and interventions. It provides a systematic way to capture what happens in a process of change, to reflect and analyse how and why it happens, and to organise and disseminate the findings (after Schouten, 2007a). It is like a diary of an intervention, trying to understand hidden mechanisms and the factors enabling or preventing change, documenting the iterative adaptation to changes and stimulating debate inside and outside the intervention team.

Process documentation starts with the hypothesis of change on which an intervention is based, looking also at the stakeholders involved. This will indicate what processes should be tracked. A four-step strategy can be used to document the processes through which changes happen.

1. Choose the tools to capture the processes of change, including such things as interviews, focus group discussions, observations, diaries, pictures and films, etc. They should be used during key-events (workshops, meetings, trainings) and key-moments (when there is tension and conflict, discussions, when solutions or compromises are found), to give a voice to key-people. Some interviews should be one-to-one to capture different perspectives and particularly to capture under represented voices.

2. Design systems to file, organise, synthesise the data collected, looking at common themes, trends and patterns and placing the findings in the larger context of the intervention and of its hypotheses of change. This can be done through articles, case-

studies, picture albums, mixed videos, portraits of people, etc. The aim is to make the information easily accessible.

3. Hold regular discussions in face-to-face or group meetings with the intervention stakeholders to analyse, understand and learn from these findings. One key aim is to learn from the documentation with a view to adjusting the project while it is still in progress.

4. Devise strategies to disseminate this information within the intervention (to learn with stakeholders) and (where appropriate) externally. Meetings, presentations and local or mass media can be used (newspapers, radio, television, internet, web 2.0). Some findings will be too sensitive to be disseminated publicly and will be for internal use only.

A process documentation specialist should coordinate these tasks. To some extent, this person should stand outside the project to retain some objectivity, although in practice he or she is nearly always a project team member. Data should be collected, compiled and analysed by all stakeholders, and potentially by outsiders such as journalists or film makers. Participation is the key to maximise learning (Schouten, 2007a and 2007b).

A workshop of practitioners and other project staff in Ghana in February 2009 concluded that the following were the most significant aims for process documentation.

- To help project staff and stakeholders to track meaningful events in the project to discern more accurately what is happening and how and why it is happening.
- To put the theory of change in touch with reality.
- To help share, disseminate and encourage debate about development processes so that organisations and individuals can take part in the learning processes.

Process documentation in Egypt, Jordan and Palestine in the EMPOWERS project

The Euro-Med Participatory Water Resources Scenarios (EMPOWERS) project introduced process documentation as a means to understand how and why changes in stakeholders capacity and participation could lead to improved water governance and management at the community, district and governorate levels in Egypt, Jordan and Palestine. In each country, a process documentation specialist worked with all stakeholders to collect information on individuals and institutions, but also cultural, traditional, economic and political patterns surrounding the intervention. Interviews with stakeholders and citizens, field visits and observations during meetings produced portraits, case-studies of successes or challenges, articles on building capacities and the resultant changes, photographs and video reports for internal analyse and for public dissemination. This new qualitative information brought a better understanding of each other's interests and responsibilities, helping to improve relationships between communities and water agencies. Within communities, water-related gender issues became less taboo. It enabled the intervention to raise public awareness of water and governance issues. A local official from Jordan said "It was a process of creating a public culture for giving, receiving and handling information, which had several individual, communal and institutional dimensions". (Abd-Alhadi et al., 2006, Schouten et al., 2007).

5.6 Qualitative information system

The *Qualitative information system (QIS)* uses participatory approaches to quantify people's qualitative perceptions, then stores, compares and analyses them for the effective and efficient planning, monitoring and adaptive assessment of WASH services. QIS is primarily based upon the Qualitative Information Appraisal (QIA), a methodology to assess the sustainability and social equity of community managed WASH services, using participatory, gender and poverty sensitive tools at community, agency and policy levels.

QIA consists of:

- the *quantified participatory assessment (QPA)*, a methodology to collect people's perceptions using participatory tools (such as welfare classifications, social mapping, pocket, matrix and rope voting, card sorting, transect walks and benefit-cost analysis), and to convert this qualitative information into numbers using indices of change, cardinal measurement and ordinal scoring (see Table 5). The key to this process is turning the qualitative assessments into scores. This needs to be a participatory exercise, validated by several groups so that the results are accepted as robust and can be compared over time and between groups.
- *Stakeholder meetings*, one-to-one or in groups to reflect on the findings of the QPA and suggest action for the intervention team, district agencies and beneficiary communities.
- *Action planning reports (APR)*, gathering the findings of the QPA and stakeholders meetings for action planning.

Options	Scores	Score given
Latrine exists but are not functional or not being used	0	
Latrine exists and is in use but they are dark, smelly and soiled with excreta	10	
Latrine exists and is in use, with adequate daylight, but soiled with excreta. No water soap or ash for hand washing with easy reach.	25	
Benchmark: Latrines are clean (no excreta in pans, walls or floor) and protected against misuse (e.g., locked after school hours)	50	
<i>In addition</i> , there is water, soap or ash for hand washing within easy reach of the children	75	
Ideal: <i>In addition</i> , Latrines are child friendly (e.g., pans are smaller, colourful walls, etc.)	100	
Reason for score		

Table 7. Scoring options to assess school latrines management (Postma et al., 2004).

QIS was developed so that information from QIAs can be compared. A computerised



Figure 17. QIS, a system to learn and improve WASH interventions' management (Casella et al., 2007).

database enables implementers to track the progress of specific communities over time and to compare different communities, to understand the factors for success and failure in the equitable and sustainable management of WASH services.

QIS can be tailored to suit local conditions and needs, yet generate uniform and comparable qualitative information. It is a powerful monitoring tool to quantify perceptions of community men and women and to monitor project progress. QIS gives district level staff a simple, gender- and poverty-sensitive database on how their services

are regarded and on the nature of planning and management processes. It can help communities manage local services in a participatory way and generates useful, accessible data problem solving. The Village Information System, a management information resource that remains with the community, includes both a detailed community map and diagrammatic depictions of the QIA results that are updated over time in subsequent QIAs to track change and plan further corrective action. (Postma et al., 2004, Casella et al., 2007).

**Participatory assessment of change over time using QIS by TWSSP, Sri Lanka
(Deirdre C. Casella, IRC)**

The Third Water Supply and Sanitation Programme (TWSSP), implemented by the National Water and Sanitation Drainage Board (NWSDB) of the Government of Sri Lanka with local government bodies, aimed to provide or improve water supply and sanitation for a 1 million women and men in rural areas and small towns.

The programme sought to ensure sustainability and community empowerment by developing the capacity of community based organisations (CBOs) and strengthening their links with local government.

TWSSP conducted a participatory assessment using QIS to monitor indicators such as how services were used and sustained, demand-responsiveness, sharing of burdens and benefits, user participation, and institutional and policy support.

The results were discussed in stakeholder meetings and action plans for improvements were defined at district and community levels. The assessment was repeated in 2006 with 25% of the original 104 sub-projects. Emphasis was given to changes that took place following the action plans and new indicators were formulated to measure outcomes of a hygiene awareness raising programme.

Marked improvements were observed in areas that had received significant attention in the action plans, including involvement of the poorest women and men, construction and use of hygienic toilets, skills for construction and maintenance of systems and, awareness and practice of hygienic behaviours such as hand washing.

Other areas continued to lag, notably the lack of CBO capacity for financial management, fault reporting and repairs and spare part management. Links with support institutions remained ad hoc and weak. Despite achievements in the construction of single pit pour flush latrines, the relatively high cost of this option effectively excluded a large number of the poorer households.

The use of the QIS tools meant that the assessment not only identified what areas were not attaining the intended results, but also captured the underlying reasons. New action plans were made as a result.

6. Conclusion and the way forward: Mixed evaluation and assessment methods

“We are here to do;
And through doing to learn;
And through learning to know;
And through knowing to experience wonder;
And through wonder to attain wisdom;
And through wisdom to find simplicity;
And through simplicity to give attention;
And through attention to see what needs to be done.”

Ben Hei Hei

6.1 Conclusion

Over recent decades, steadily increasing attention has been paid to the results of development interventions. This, together with an increasing demand for development agencies to be transparent and accountable, has led to a focus on impacts and to quantifying them through impact evaluations. Some however argue that evaluations should not only focus on proving retrospectively what has changed as a result of an intervention, but also on improving ongoing interventions, through assessing how and why they lead to changes.

Development interventions can be modelled in different ways, each of which has methodological implications on their evaluation. They can be seen as simple systemic chains of steps producing predictable impacts, which can be quantified through evidence-based evaluations to determine ‘what works and what does not’. They can also be understood as multiple mechanisms leading to different outcomes depending on the context, which can be understood through theory-based and realistic evaluations to find out ‘what works where and when’. Finally, some see them as complex, dynamic and adaptive systems from which unpredictable outcomes and impacts emerge, which should be analysed with complexity evaluations, to assess ‘how things work and why’.

The WASH sector and its governance, composed of numerous interrelated and dynamic stakeholders and influencing factors, can also be seen as complex adaptive systems. This does not mean that all WASH interventions are necessarily complex, but the complexity of the sector and its governance has to be taken into account when designing, implementing assessing and evaluating them.

With the growing interest in improving impact evaluation and assessment of development interventions, some currently strongly advocate evidence-based methods, especially randomised controlled trials, as the most rigorous and scientific evaluation method. RCTs clearly can be used to evaluate certain types of WASH interventions, especially those

focussing on the household level, but cannot be considered as the gold standard in all cases (EES, 2007). They are recognised as having various weaknesses and as rarely being appropriate in complex situations where outcomes are the results of multiple factors interacting simultaneously (EES, 2007 and NONIE, 2008), which is often the case in the WASH sector. The strengths of other quantitative methods, such as quasi-experimental evaluations, are also recognised but these too are not appropriate for all types of interventions (NONIE, 2008). Other evaluation methods are now also advocated for use in the development and WASH sectors, including theory-based and realistic evaluations (NONIE, 2008), complexity evaluations (Ramalingam, 2008 and Hospes, 2008) and change assessments (Earl et al., 2001a, Narayan, 1993). Table 8 (following page) summarises all the evaluation and assessment methods presented in this Thematic Overview Paper, as well as their main assets, drawbacks and challenges.

Quantitative and qualitative, evidence-based, realistic, complexity and change methods all have strengths and limitations, and can all contribute to prove and improve WASH interventions. However, because of the complexity of the sector and the diversity of its interventions (with different objectives, contexts, mechanisms, outcomes, impacts), and because numerous stakeholders require different information, presented in different ways, single assessment or evaluation methods are not sufficient. It is increasingly recognised that impact evaluation and assessment is complex, particularly for multi-dimensional interventions such as capacity-building or sectoral development, and for governance issues. These require a variety of different methods that can take into account this inherent complexity (EES, 2007). Both quantitative and qualitative data should be used, with the balance between the two being determined by the questions being addressed, the type of intervention and the evaluation design (NONIE, 2008), the rigour being determined by the match between the methods and an interventions' theory of change. No example of such a multi-method evaluation and assessment was found as part of this study. RiPPLE (mentioned in 5.4 above) did try to use a variety of tools, but essentially qualitative ones (Most Significant Change, Outcome Mapping, Social Network Analysis, interviews and portraits).

There is therefore a need to shift from single to multiple evaluation and assessment methods, and from only proving objectives to also improving interventions. Mixed evaluation and assessment methods appear to be the best way to learn more about the WASH sector, its governance and their complexity, to understand not only what changes WASH interventions can bring, but also how and why these changes occur, and how can they be optimised and sustained. This shift might take time, due to the current preferences of funding and implementing agencies, the current practice of evaluation experts, and the potential cost implications of using mixed methods. Before attempting to change others, we might have to change ourselves.

6.2 Summary table of all the evaluation and assessment methods above

Table 8. Summary table of all the evaluation and assessment methods presented in this Thematic Overview Paper

	Methods and tools	Monitoring	Evaluation	Assessment	Brief description	Main assets	Main drawbacks and challenges	Main sources (see References and TOP books)
Evidence-based evaluations	Longitudinal studies		X		Study of the same population through repeated observations before and after a WASH intervention, or in the long term.	<ul style="list-style-type: none"> • Capture long-term effects, excluding time-invariant unobserved individual differences • Can be used in 'at scale' projects in which a counterfactual cannot be studied 	<ul style="list-style-type: none"> • Do not compare beneficiaries and counterfactual • Cannot always attribute changes to an intervention • Do not always detect confounding effects • Long and expensive to implement 	World Bank, 2009
	Case-control studies		X		Retrospective study comparing <i>cases</i> (e.g. with health symptoms) and <i>controls</i> (e.g. without health symptoms), looking back at their initial WASH differences to link them with symptoms.	<ul style="list-style-type: none"> • Attempts to compare WASH beneficiaries to a counterfactual • Cheaper and quicker to implement than Longitudinal studies 	<ul style="list-style-type: none"> • Short term studies that cannot capture long-term effects • Counterfactual (controls) can be biased • Do not always take confounders into account 	Briscoe et al., 1985 Hunter et al., 2002
	Randomised controlled trials (RCTs)		X		Comparison of random beneficiaries (<i>treatment group</i>) with a random counterfactual	<ul style="list-style-type: none"> • Randomisation theoretically makes the treatment and comparison groups statistically 	<ul style="list-style-type: none"> • Randomisation is often difficult to implement • Require highly skilled and 	Baker, 2000 Ravallion, 1999 Ravallion, 2008

				(<i>control group</i>).	<p>identical (i.e. unbiased) and can separate impact from confounders</p> <ul style="list-style-type: none"> • Impact is well identified, quantified and attributed to the intervention 	<p>experienced evaluation experts</p> <ul style="list-style-type: none"> • Can be applied to a range of simple and complicated intervention but not to complex ones • Depending on the design, the reasons behind impact are not always fully explored 	
	Quasi-experimental studies		X	Non-randomised comparison of beneficiaries (<i>treatment group</i>) with a statistically constructed counterfactual (<i>comparison group</i>).	<ul style="list-style-type: none"> • Can be applied to a wider range of interventions as there is no randomisation • Impact is well identified, quantified and attributed to the intervention 	<ul style="list-style-type: none"> • Require highly skilled and experienced evaluation experts • Counterfactual might be affected by unknown biases • Can be applied to a range of simple and complicated intervention but not to complex ones • Depending on the design, the reasons behind impact are not always fully explored 	
	Non-experimental studies		X	Use of secondary data and statistical methods to evaluate impact of a WASH intervention.	<ul style="list-style-type: none"> • Usually cheaper than RCTs and quasi-experimental evaluations • Less difficult to implement as they rely on secondary data 	<ul style="list-style-type: none"> • Require highly skilled and experienced evaluation experts • Do not compare beneficiaries and counterfactual 	World Bank, 2009

							<ul style="list-style-type: none"> • Might be less effective to precisely quantify impact, depending on the data available 	
Realistic evaluations	Theory-based evaluations	X	X		Use of a programme theory to determine and test different causal paths along the causal chain	<ul style="list-style-type: none"> • Can explore multiple causal paths, alternative and simultaneous ones • Include the context of interventions 	<ul style="list-style-type: none"> • A programme theory, even detailed, cannot fully capture a complex intervention • Was never applied to a WASH intervention 	Rogers, 2008
	Realistic evaluations	X	X		Use of Context-Mechanism-Outcome configurations to test theories along the causal chain	<ul style="list-style-type: none"> • Explores what works or not, in which context and through which mechanisms, at all steps of an intervention 	<ul style="list-style-type: none"> • Was never applied to development interventions 	Pawson et al., 1997
	Minimum Evaluation Procedures	X	X		Series of evaluations at various stages of a WASH intervention, to determine how WASH systems work, how they are used and managed, as well as their impact	<ul style="list-style-type: none"> • Pragmatic, quantitative and qualitative field oriented evaluation method looking at the various links of the intervention causal chain • Method focusing on health, but also socio-economic changes 	<ul style="list-style-type: none"> • Apparently not used anymore since the 1990s 	WHO, 1983 Schultzberg et al., 1983
	Hygiene Evaluation Procedures	X	X		Participatory, qualitative studies of hygiene behaviours and changes	<ul style="list-style-type: none"> • Focuses on the intermediate links of the causal chain, to attribute hygiene changes to the WASH intervention • Can be used to evaluate health impacts, but also to understand 	<ul style="list-style-type: none"> • Focuses on hygiene and health only • Does not quantify health impacts but assesses hygiene changes that can lead (or not) to them 	Almedom et al., 1997 Cairncross et al., 1994

						certain underlying social processes		
Change assessments	Balanced scorecard	X		X	Method to assess public services agencies through few key objectives in four perspectives of performance that counterbalance one another.	<ul style="list-style-type: none"> • Provides a broad picture of a public agency's performance and sustainability • Shows strengths and weaknesses, and encourages managers to optimise the performance of their agency by balancing the four perspectives over the longer term 	<ul style="list-style-type: none"> • Does not evaluate a specific intervention, but assesses an agency • Does not give clear indications of how to proceed to improve performance and sustainability. Managers must decide and balance changes 	Estis, 1998
	Most Significant Change (MSC)	X	X	X	Participatory assessment and selection of changes considered as most important by various stakeholders, at different levels of an intervention.	<ul style="list-style-type: none"> • Apparently very simple to implement • Fully qualitative, does not require pre-defined indicators, and therefore very flexible • Many stakeholders, including beneficiaries, can participate • Can make all stakeholders focus on change, if implemented properly 	<ul style="list-style-type: none"> • Steps as establishing and training a responsible team and 'domains of change' are critical to get useful results • Collecting and analysing stories objectively, in a real participatory manner is difficult • Stories are not necessarily representative, 'just' significant 	Davies et al., 2005 Davies, 2009
	Social network analysis	X	X	X	Tool to model and analyse social networks (structure, members, relationships), and plan and assess networks	<ul style="list-style-type: none"> • Useful to analyse any network, especially large, complex ones • Can identify key partners, strong ones and ones to be 	<ul style="list-style-type: none"> • The choice of the information used to model and analyse a network is critical. Different choices can lead to different 	Davies, 2008 Hanneman et al., 2005

					change strategies.	strengthened <ul style="list-style-type: none"> • Can be used to find adequate 'network pathways' to influence specific stakeholders through others 	results <ul style="list-style-type: none"> • Proper stakeholder participation is often challenging • Can be complex to implement and analyse in large networks 	
Outcome mapping	X	X	X		Method to plan the outcomes (changes) a complex intervention and its boundary partners want to influence; and to monitor, analyse and assess the progress towards them.	<ul style="list-style-type: none"> • Flexible method, where the desired changes, progress markers and monitoring/ analysis/ assessment strategies can be readapted when a project changes or when new outcomes emerge • This method is particularly well adapted for large, complex projects 	<ul style="list-style-type: none"> • Requires adequate training • Requires lot of documentation of processes, in a participatory manner, which is demanding and challenging • Does not attribute impacts to an intervention, but attributes changes to the influence of an intervention 	Earl et al., 2001a Earl et al., 2001b
Process documentation	X	X	X		Systematic way to capture what happens in a process of change, to reflect and analyse how and why it happens, and to organise and disseminate the findings	<ul style="list-style-type: none"> • Very flexible tool, offering a large variety of ways to document change and disseminate the findings • Can help project staff, partners, beneficiaries as well a wider public better understand the process of change, how and why it happens 	<ul style="list-style-type: none"> • Defining hypotheses of change is critical for the success of process documentation • Requires a dedicated and trained staff with specific skills, working 'inside' the project, but also taking a step back to remain critical 	Schouten, 2007a Schouten, 2007b

							<ul style="list-style-type: none"> • Some findings might be too sensitive to be disseminated 	
	Qualitative Information System	X	X	X	Participatory approach to quantify people's qualitative perceptions; then to store, compare and analyse them for the effective and efficient planning, monitoring and adaptive assessment of WASH services	<ul style="list-style-type: none"> • Enables to compare and assess qualitative information over space and time by quantifying it • Gives indication, based on participation, on the reasons of success or failure, and how to adapt or correct the project and its management when necessary 	<ul style="list-style-type: none"> • Based on a variety of tools requiring skills and experience • The choice of the methodology to score qualitative data is critical. Different methodologies might lead to different results. 	<p>Postma et al., 2004</p> <p>Casella et al., 2007</p>

7. Glossary and acronyms

7.1 Acronyms

ADRA	Adventist Development and Relief Agency
AMREF	African Medical and Research Foundation
APR	Action Planning Reports
BP	Boundary Partners
BSC	Balanced Scorecard
CBO	Community-Based Organisation
CMO	Context-Mechanism-Outcome
EMPOWERS	Euro-Med Participatory Water Resources Scenarios
FLoWS	Forum for Learning on Water and Sanitation
GTZ	Deutsche Gesellschaft für Technische Zusammenarbeit
HEP	Hygiene Evaluation Procedures
IDWSSD	International Drinking Water Supply and Sanitation Decade
IRC	IRC International Water and Sanitation Centre
JMP	Joint Monitoring Programme
KCPP	Katine Community Partnerships Programme
LPA	Learning and Practice Alliance
M&E	Monitoring and Evaluation
MDGs	Millennium Development Goals
MEP	Minimum Evaluation Procedure
MPA	Methodology for Participatory Assessment
MSC	Most Significant Change
NGO	Non-Governmental Organisation
PHAST	Participatory Hygiene And Sanitation Transformation
OM	Outcome Mapping
PRA	Participatory Rural Appraisal
QIS	Qualitative Information System
RIPPLE	Research-inspired Policy and Practice Learning in Ethiopia and the Nile Region

RBM	Results-Based Management
RCT	Randomised Controlled Trial
QIA	Qualitative Information Appraisal
QIS	Qualitative Information Systems
QPA	Quantified Participatory Assessment
SNA	Social Network Analysis
UNICEF	United Nations Children's Fund
TWSSP	Third Water Supply and Sanitation Programme
WASH	Water, Sanitation and Hygiene
WHO	World Health Organization

7.2 Glossary

This glossary gives definition of all terms highlighted in italics in the above text. Definitions have been taken or adapted from various sources, which can be found in the references.

Accountability	<p>Obligation to demonstrate that work has been conducted in compliance with agreed rules and standards or to report fairly and accurately on performance results vis-à-vis mandated roles and/or plans. This may require a careful, even legally defensible, demonstration that the work is consistent with the contract terms.</p> <p>Accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources. For evaluators, it connotes the responsibility to provide accurate, fair and credible monitoring reports and performance assessments. For public sector managers and policy-makers, accountability is to taxpayers / citizens (DAC, 2002).</p> <p>Accountability to service users is an issue often overlooked in service delivery.</p>
Action Planning Reports	<p>In the Qualitative Information System / Qualitative Information Appraisal (QIS/QIA) approach, APR present findings of the Quantified Participatory Assessment (QPA) and suggestions from stakeholder meetings in a manner most suitable for action planning by project management and project communities (Potsma et al., 2004).</p>
Activities	<p>Actions or work through which inputs, such as funds, technical</p>

assistance and other types of resources are mobilized to produce specific outputs (DAC, 2002).

Adaptive agents	Individual, teams and organisations are adaptive agents. They perceive the complex system(s) around them and act on these perceptions. Their view of the world dynamically influences, and is influenced by, events and changes within system(s) (Ramalingam et al., 2008).
Alternative causal strands	In theory-based evaluations, there are several causal paths through which an intervention can work. In many case, these alternative causal strands are effective in particular context (after Rogers, 2008).
Assessment	On-going, participatory investigation of the processes of change affecting an intervention, to learn why and how changes occur, with the objective of improving the assessed intervention as well as future ones (after Parker et al., 2001).
Attribution	<p>The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.</p> <p>Note: Attribution represents the extent to which observed development effects can be attributed to a specific intervention or to the performance of one or more partner taking account of other interventions, (anticipated or unanticipated) confounding factors, or external shocks (DAC, 2002).</p> <p>Related term: causality.</p>
Balanced Scorecard	Quantitative and qualitative method to assess public services. It involves identifying and measuring a few key objectives in four perspectives of performance that counterbalance one another, to draw a balanced picture of a public agency's performance and sustainability.
Beneficiaries	Individuals, groups, or organisations, whether targeted or not, that benefit, directly or indirectly, from a development intervention (DAC, 2002).
Black box evaluation	Evaluation which gives a finding on impact, but no indication as to why the intervention is or is not working. Answering the why question requires looking inside the box, or along the results chain (OECD, 2006).
Black boxes	In realistic evaluations, black boxes represent an organisation, social interaction or a social force field in which input is converted

(realistic evaluation)	into output, output into income, and outcome into impact, but not necessarily through a linear connection (Hospes, 2008)
Boundary partners	Individuals, groups, or organisations with whom the programme interacts directly and with whom the programme can anticipate opportunities for influence. These actors are called boundary partners because, even though the programme works with them to effect change, it does not control them. The power to influence development rests with them (Earl, 2001a).
Case-control study	Retrospective study of events that preceded the onset of disease (e.g. diarrhoea) in a group of individuals. In a case-control study, hypotheses are tested by comparing the incidence of a preceding event in those with disease (cases) with a group of individuals who do not appear to have disease (controls) (Hunter et al., 2002).
Cases	Group of people having a disease (e.g. diarrhoea), compared to controls in case-control studies (Hunter et al., 2002).
Causal chain	Synonym of results chain.
Causality	The fact that a step from the causal chain or programme theory results in the following step. The second step can then be attributed to the first one. Related term: Attribution.
Causal link	Link of cause and effect between two steps of the causal chain or programme theory of an intervention.
Causal path	The way the intervention moves from one link to another in the causal chain, or in the programme theory.
CMOs configurations	In realistic evaluations, context-mechanism-outcome configurations are hypotheses made to test how interventions move along the different steps of a complicated programme theory. They can be tested through logical thinking and surveys to determine which configurations work or not, and if so, where and when.
Comparison group	Group of people used in quasi-experimental evaluations to identify the counterfactual of what would have happened without the intervention. The comparison group is designed to be representative of the treatment group of participants with one key difference: the comparison group did not participate to the intervention (Ravallion, 1999)

	Related term: Control group (used in experimental evaluations only) (World Bank, 2009).
Complexity evaluation	Evaluation tracking unpredictable, emerging changes or outcomes throughout implementation of complex interventions or interventions implemented in dynamic, complex adaptive systems (after Hospes, 2008 and Rogers, 2008).
Complexity science	A body of interdisciplinary knowledge about the structure, behaviour and dynamics of change in complex systems - open evolutionary systems with multiple, strongly interrelated components - including <i>complex adaptive systems</i> - complex systems where the components are self-organising and dynamic (after Sanders, 2003).
Complex adaptive system	Open evolutionary systems in which the components are strongly interrelated, self-organising and dynamic (Sanders, 2003).
Comprehensive evaluation	Evaluation including needs analysis, ex-ante impact evaluation, monitoring, process evaluation, cost-effectiveness or cost-benefit evaluation, and impact evaluation (after Baker, 2000).
Confounders	Factors or events that are correlated with the outcomes but are not caused by the intervention. Their effect on outcomes needs to be taken into account, so it is not attributed it to the intervention (Poulos et al., 2006).
Confounding factors	Synonym: Confounders.
Contagion effect	Contagion occurs when a project is initiated by another agency in the comparison area, or when changes in the intervention area influence behaviour change in the comparison area. The evaluation design should include data collection to capture this effect (after IEG, 2008). Synonym: Contamination effect.
Contamination effect	Synonym: Contagion (White, 2006).
Context	In realistic evaluations, context refers to the particular implementation environment or characteristics of participants within which a specific mechanism can work.
Control group	Group of people not included in an intervention and used in

	<p>experimental evaluations to identify the counterfactual of what would have happened without the intervention (Ravallion, 1999).</p> <p>Related term: Comparison group (used in quasi-experimental evaluations only) (World Bank, 2009).</p>
Controls	Group of people having no disease, compared to cases in case-control studies (Hunter et al., 2002).
Correlation	A statistical measure of the degree of relationship between or among variables (Sanders et al., 1994).
Counterfactual	The situation or condition which hypothetically may prevail for individuals, organisations, or groups were there no development intervention (DAC, 2002).
Cost-benefit analysis	A monetary measure of all the cost and benefits of an intervention, including health and time savings ones, that can be used to compare various interventions and choose the one that yields more benefits (after Baker, 2000 and Belli et al., 1997).
Cost effectiveness analysis	A monetary measure of the scale of impacts of an intervention compared to its cost, that can be used to compare various interventions with the same impact effectiveness and choose the cheapest (Baker, 2000 and Belli et al., 1997).
Customer perspective	One of the four perspectives of the Balanced Scorecard adapted for the public sector, assessing whether a government agency serves its customers and how satisfied they are (Estis, 1998).
Diagram (in network analysis)	<p>In network analysis, visual representation of a network, giving information on structure, its members, their attributes as well as the nature of their relationships (after Davies, 2008).</p> <p>Related term: Matrix.</p>
Difference in difference method	Method used in experimental or quasi-experimental evaluations, comparing a treatment and comparison group (first difference), before and after a programme (second difference) (Ravallion, 1999).
Double difference method	Synonym of difference in difference method (Ravallion, 1999).
Emergence	When the specific outcomes, and the means to achieve them, emerge during implementation of an intervention (Rogers, 2008).

Environmental study	Sometimes used to refer to a non-experimental (evidence-based) study.
Evaluation	<p>The systematic and objective investigation of the worth or merit of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons into the decision-making process of both recipients and donors (after DAC, 2002 and Sanders, 1994).</p> <p>Related term: Comprehensive evaluation.</p>
Evaluation planning	The third stage of outcome mapping. It helps the programme identify evaluation priorities and develop an evaluation plan (Earl et al., 2001a).
Evidence-based evaluation	Evaluation aiming to find measurable changes that can be directly attributed to specific policies or interventions, using longitudinal or case-control studies, or experimental, quasi-experimental or non-experimental evaluations (after Hospes, 2008).
Experimental evaluation	<p>Evaluation gathering a set of individuals (or other unit of analysis) equally eligible and willing to participate in the programme and randomly dividing them into two groups: those who receive the intervention (treatment group) and those from whom the intervention is withheld (control group, used as counterfactual). The impacts can be estimated after the intervention by comparing the average values of indicators of the two groups (World Bank, 2009).</p> <p>Synonym: Randomised controlled trial.</p>
Formative evaluation	<p>Evaluation intended to improve performance, most often conducted during the implementation phase of projects or programmes.</p> <p>Synonym: Process evaluation (DAC, 2002).</p>
Goal	In the logical framework matrix, the higher-order objective to which a development intervention is intended to contribute (DAC, 2002).
Hypothesis of change	Hypotheses of the change that an intervention wants to trigger or to achieve.
Impacts	Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly,

	intended or unintended (DAC, 2002).
Impact evaluation	Evaluation intended to determine whether an intervention had the desired effects on individuals, households, and institutions and whether those effects are attributable to the intervention. Impact evaluations can also explore unintended, positive or negative, consequences on beneficiaries (Baker, 2000).
Impact heterogeneity	Variability of the impact of an intervention according to its design, context, and beneficiary characteristics. If it is to be policy relevant, an impact study must maintain a balance between isolating the specific context which allows the impact and the ability to generalize (IEG, 2008).
Indicator	Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor (DAC, 2002).
Inputs	The resources and activities of a development intervention (Poulos et al., 2006).
Instrumental variables	In evidence-based evaluations, variables that matter to participation, but not to outcomes given participation. If such variables exist then they identify a source of exogenous variation in outcomes attributable to the programme - recognising that its placement is not random but purposive. The instrumental variables are first used to predict programme participation, then one sees how the outcome indicator varies with the predicted values (Ravallion, 1999).
Intentional design	The planning stage of outcome mapping, where a programme reaches consensus on the macro level changes it would like to help bring about and plans strategies to provide appropriate support (Earl et al., 2001a).
Logical framework	Management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention (DAC, 2002).
Longitudinal study	A correlational, observational research study that involves repeated observations of the same items over long periods of time,

	sometimes several decades. Longitudinal studies are often used in medicine to uncover predictors of certain diseases (After Wikipedia, 2008).
Matching	Method used in quasi-experimental evaluations to pick an ideal comparison group from a larger survey. The comparison group is matched to the treatment group on the basis of a set of observed characteristics, or using the propensity score. A good comparison group comes from the same economic environment and was administered the same questionnaire by similarly trained interviewers as the treatment group (Ravallion, 1999). Related term: Propensity score.
Matrix (in network analysis)	In network analysis, tabular representation of a network, giving information on its structure, its members, their attributes as well as the nature of their relationships (after Davies, 2008). Related term: Diagram.
Mechanism	In realistic evaluations, the mechanism is the precise way in which an intervention works within a given context to produce a particular outcome (Gill et al., undated).
Meta-analysis	The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators (DAC, 2002).
Mission perspective	One of the four perspectives of the balanced scorecard adapted for the public sector, assessing whether a government agency accomplishes the primary public service mission for which it has been created (Estis, 1998).
Monitoring	A continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds (DAC, 2002).
Most significant change	Participatory change assessment method, consisting in regularly collecting stories of significant change from intervention stakeholders, analysing them and selecting the most significant one. This process can be followed at several levels of an intervention (e.g. community, district, country) (Davis et al., 2005).
Non experimental	An evaluation design that can be used when it is not possible to

evaluation	randomly select a control group, identify a suitable comparison group through matching methods or use reflexive comparisons. In such situations, participants can be compared to non-participants using statistical methods such as instrumental variables and regression to account for differences between the two groups (World Bank, 2009).
Objectives	Synonym of purpose, or outcomes (in the logical framework) (NZAID, 2002).
Operational efficiency	One of the four perspectives of the balanced scorecard adapted for the public sector, assessing whether a government agency delivers its public services at the lowest possible cost (Estis, 1998).
Organisational learning and improvement	One of the four perspectives of the balanced scorecard adapted for the public sector, assessing the whether a government agency makes the capital, labour and processes investments it needs to prepare its future and ensure its sustainability (Estis, 1998).
Outcome mapping	An assessment method focusing on outcomes, i.e. the behavioural changes of its boundary partners. It consist of three phases: intentional design (to list the boundary partners, the targeted outcomes, design a strategy to observe them and create progress markers), outcome and performance monitoring (using journals to monitor performance, strategy and outcomes), and evaluation planning (to use findings to design an evaluation plan) (Earl et al., 2001).
Outcome and performance monitoring	The second stage of outcome mapping. It provides a framework for the ongoing monitoring of the programme's actions in support of the outcomes and the boundary partners' progress towards the achievement of outcomes. It is based largely on systematised self-assessment (Earl et al., 2001a).
Outcome journal	In outcome mapping, data collection tool for monitoring the progress of a boundary partner in achieving progress markers over time (Earl et al., 2001a).
Outcomes	Changes in the behaviour, relationships, activities, or actions of the people, groups, and organisations with whom a programme works directly, that can be logically linked to a programme's activities, although not necessarily directly caused by them (Earl, 2001a).

Outputs	The products, capital, goods and services which result directly from the intervention activities (Poulos et al., 2006).
Performance journal	In outcome mapping, a data collection tool for monitoring how well the programme is carrying out its organisational practices (Earl et al., 2001a).
Placebo	A treatment without intrinsic therapeutic value, but administered as if it were a therapy, either in medical treatment or in clinical trials (Wikipedia, 2009).
Process documentation	Systematic way to capture what happens in a process of change and how it happens, to reflect and analyse why it happens, and to organise and disseminate the findings (after Schouten, 2007a).
Process evaluation	An evaluation of the internal dynamics of implementing organisations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these (DAC, 2002). Synonym: Formative evaluation.
Programme logic	Synonym: Programme Theory (Rogers, 2008).
Programme theory	A variety of ways of developing a causal model linking programme inputs and activities to (a) chain(s) of intended or observed outcomes, to guide theory-based evaluations (Rogers, 2008). Synonyms: Programme logic, Theory of change.
Progress markers	In outcome mapping, a set of graduated indicators of changed behaviours for a boundary partner that focus on the depth or quality of change (Earl et al., 2001a).
Propensity score	Statistical matching method used in quasi-experimental evaluations to predict the probability of participation [to an intervention] given observed characteristics. The closer the propensity score, the better the match between the treatment and comparison groups (Ravallion, 1999). Related term: Matching.
Purpose	Synonyms: Objectives, Outcomes (in the logical framework) (NZAID, 2007).
Qualitative information	In the Qualitative information system (QIS) approach, QIA is an assessment method using the quantified participatory assessment (QPA), stakeholder meetings and action planning reports (APR)

appraisal	(Postma et al., 2004).
Qualitative information system	Assessment tool using the qualitative information appraisal tool to collect and quantify qualitative information at regular intervals, and storing this information on a computer database to facilitate analysis of comparative progress across time and locations (Postma et al., 2004).
Quantified participatory assessment	In the qualitative information system / qualitative information appraisal (QIS/QIA) approach, QPA is an assessment tool using participatory methods to generate people's perceptions, and convert this information into numbers using indices of change, cardinal measurement and ordinal scoring (Postma et al., 2004).
Quasi-experimental evaluation	Evaluation comparing a treatment group (that receives the intervention) and a comparison group (that does not), matching both groups to address the selection bias using statistical methods rather than randomisation. These methods model the selection process and so control for these variables in the analysis of outcomes (after World Bank, 2009, OECD, 2006).
Randomised controlled trial	Synonym: Experimental evaluation (World Bank, 2009).
Realistic evaluation	Evaluation using context-mechanism-outcome configurations to analyse how an intervention is received under certain social and cultural conditions (context) and how it triggers (mechanism) a response (outcome) from people (Hospes, 2008).
Recursive causality	When the cause and effect relationships in the programme theory (theory-based evaluations) of a complex intervention are mutual, multi-directional and multilateral, rather than unidirectional like in the causal chain (after Patton, 1997, quoted in Rogers, 2008).
Reflexive comparison	Evaluation in which a baseline survey of participants is done before the intervention, and a follow-up survey done after. The baseline provides the comparison group, and impact is measured by the change in outcome indicators before and after the intervention (Ravallion, 1999).
Resources	The internal or external financial, human, social, institutional and material resources supporting or constraining a development intervention (Poulos et al., 2006).
Results	In the causal chain, results are the output, outcome or impact (intended or unintended, positive and/or negative) of a

	development intervention (DAC, 2002).
Results chain	<p>The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts, and feedback (DAC, 2002).</p> <p>Synonym: Causal chain.</p>
Selection bias	<p>In evidence-based evaluations, when an intervention is not allocated randomly, the treatment group is selected (or self select) on account of special characteristics, observable or not. The comparison group needs to be selected in such a way that it has similar characteristics so that the evaluation does not yield biased results. This can be done through matching (World Bank, 2009, IEG, 2008).</p>
Simultaneous causal strands	<p>In theory-based evaluations, two or more simultaneous strands in the programme theory that are all required for the intervention to succeed (Rogers, 2008).</p>
Single difference method	<p>In evidence-based evaluations, method measuring the difference in the output or outcome either (1) before versus after the intervention, or (2) between project and comparison groups. Before versus after is not a good impact measure as it fails to control for other factors. The single difference project versus comparison groups fails to allow for differences between the two groups which may have existed prior to the intervention. The double difference method takes care of these two problems (OECD, 2006).</p> <p>Related term: Double difference method.</p>
Social network analysis	<p>Method to model and analyse networks, their structure, their members and the nature of their relationships. It can be used to describe existing networks, plan changes to improve or enlarge them, and to assess these changes (Davies, 2008).</p>
Spillover effect	<p>In evidence-based evaluations, spillover occurs when members of the comparison or control group experience welfare impacts, either positive or negative, from the project. The evaluation design should include data collection to capture this effect (IEG, 2008).</p> <p>Related term: Contamination or contagion effect.</p>
Stakeholder meetings	<p>In the qualitative information system / qualitative information appraisal (QIS/QIA) approach, stakeholders meetings are</p>

	individual or group meetings which use the findings from the quantified participatory assessment (QPA) to probe factors underlying the performance reflected in the scores, and to suggest action points for both project management and project communities (Postma et al., 2004).
Strategy journal	In outcome mapping, data collection tool for monitoring the strategies a programme uses to encourage change in the boundary partner (Earl et al., 2001).
Summative evaluation	A study conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of the programme (DAC, 2002). Synonym: Impact evaluation.
Theory-based evaluation	Theory-based evaluations focus on unpacking the theoretical or logical sequence(s) (programme theory, or mechanisms) by which an intervention is expected to bring about its desired effects (Cabinet Office, 2003).
Theory of change	Synonym: Programme theory (Rogers, 2008).
Treatment group	In evidence-based evaluations, those who receive the intervention, and who will be compared to the control group (experimental evaluations) or the comparison group (quasi-experimental evaluations) (World Bank, 2009).
Triangulation	The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an evaluation or assessment. By combining multiple data sources, methods, analyses or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observer or single theory studies (DAC, 2002).
Water governance	The range of political, social, economic and administrative systems that are in place to develop and manage water resources, and the delivery of water services, at different levels of society (Rogers et al., 2003).

8. TOP books, articles, papers

8.1 Evidence-based, theory-based and realistic evaluations

Baker, J. L. (2000). *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington, D. C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

<http://siteresources.worldbank.org/INTISPMA/Resources/handbook.pdf>

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London, UK, Sage Publications.

Rogers, P. J. (2008). 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions'. In: *Evaluation*, vol. 14, no. 1, p. 29-48.

8.2 Water Governance

Plummer, J. and Slaymaker, T. (2007). *Rethinking governance in water services*. Working Paper 284. London, UK, Overseas Development Institute. Retrieved 19 March 2009 from the World Wide Web:

<http://www.odi.org.uk/resources/odi-publications/working-papers/284-rethinking-governance-water-services.pdf>

Rogers, P. and Hall, A. W. (2003). *Effective Water Governance*. TEC background paper no 7. Global Water Partnership Technical Committee. Stockholm, Sweden, Global Water Partnership. Retrieved 19 March 2009 from the World Wide Web:

<http://www.gwpforum.org/gwp/library/TEC%207.pdf>

8.3 Complexity and development

Mowles, C.; Stacey, R. and Griffin, D. (2008). 'What contribution can insights from the complexity sciences make to the theory and practice of development management?' In: *Journal of International Development*, vol. 20, issue 6, p. 804-820.

Ramalingam, B.; Jones, H.; Reba, T. and Young, J. (2008). *Exploring the science of complexity : Ideas and implications for development and humanitarian efforts*. ODI Working Paper 285. London, UK, Overseas Development Institute. Retrieved 19 March 2009 from the World Wide Web:

<http://www.odi.org.uk/resources/odi-publications/working-papers/285-science-complexity.pdf>

8.4 Qualitative assessment tools

Catley, A.; Burns, J.; Abebe, D. and Suji, O. (2008). *Participatory Impact Assessment : A Guide for Practitioners*. Medford, Feinstein International Center, Tufts University. Retrieved 19 March 2009 from the World Wide Web:
<https://wikis.uit.tufts.edu/confluence/display/FIC/Participatory+Impact+Assessment--+a+Guide+for+Practitioners>

Davies, R. and Dart, J. (2005). *The Most Significant Change (MSC) Technique*. Retrieved 19 March 2009 from the World Wide Web:
<http://www.mande.co.uk/docs/MSCGuide.pdf>

Davies, R. (2009). *Rick's methods*. This blog presents various methods to monitor and evaluate and assess change in development projects. Retrieved 19 March 2009 from the World Wide Web:
<http://mande.co.uk/special-issues/>

Douthwaite, B.; Kuby, T.; Van de Fliert, E. and Schulz, S. (2003) 'Impact pathway evaluation : an approach for achievement and attributing impact in complex systems'. In: *Agricultural Systems*, n° 78, p. 243-265. Retrieved 15 May 2009 from the World Wide Web:
http://boru.pbwiki.com/f/ag_syst_IPE.pdf

Earl, S.; Carden, F. and Smutylo, T. (2001a). *Outcome Mapping, Building Learning and Reflection into Development Programs*. Ottawa, Canada, International Development Research Centre. Retrieved 19 March 2009 from the World Wide Web:
<http://www.idrc.ca/openebooks/959-3/>

Guijt, I. (2008). *Critical Reading on Assessing and Learning for Social Change : A Review*. Brighton, UK, Institute of Development Studies. Retrieved 12 May 2009 from the World Wide Web:
http://www.justassociates.org/projects_files/Db21web.pdf

Hanneman, R. A. and Riddle, M. (2005). *Introduction to social network methods*. Retrieved 19 March 2009 from the World Wide Web:
<http://faculty.ucr.edu/~hanneman/nettext/>

Narayan, D. (1993). *Participatory evaluation : Tools for managing change in water and sanitation*. World Bank Technical Paper 207. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:
http://books.google.fr/books?id=SA_CG4ygWuQC&dq=narayan+managing+change&prints_ec=frontcover&source=bl&ots=kqv3iLgm9F&sig=NGLWxxhzigKqU847XIZW0jEhxKvl&hl=fr&ei=2Sq2SYH5LITR-Aaa37iICw&sa=X&oi=book_result&resnum=1&ct=result

Postma, L.; James, A. J. and Van Wijk C. (2004). 'QIS : A New Participatory Management Tool to Assess and Act on Field Reality'. Paper presented at the 30th WEDC International Conference, Vientiane, Lao PDR. Retrieved 19 March 2009 from the World Wide Web:
<http://www.irc.nl/page/37607>

Schouten, T. (2007). *Learning Alliance Briefing No 6: Process Documentation*. Briefing note prepared for a SWITCH training workshop on process documentation, Lodz, Poland, 1-5 July 2007. Retrieved 19 March 2009 from the World Wide Web:
http://www.switchtraining.eu/fileadmin/template/projects/switch_training/db/event_upload_folder/34/Briefing_Note_6_-_Process_documentation.pdf

9. TOP websites

9.1 International Evaluation Networks

International Development Evaluation Association (IDEAS)

<http://www.ideas-int.org/>

IDEAS is a global network of development practitioners and evaluators committed to capacity building, networking, applying innovative methodological approaches, and sharing knowledge, especially in developing countries and countries in transition. Its website contains articles, reports, case-studies, book reviews and other documents on monitoring and evaluation for development.

International Initiative for Impact Evaluation (3ie)

<http://www.3ieimpact.org/>

3ie was established in 2008 on a recommendation from the Evaluation Gap Working Group from the Centre for Global Development (see below). 3ie's Members include government officials from developing countries who have a strong interest in issues of effectiveness and accountability, as well as representatives of bilateral donor agencies, multilateral agencies, non-governmental organizations, and foundations. Its declared aim is to improve the lives of poor people in low- and middle-income countries by providing, and summarizing, evidence of what works, when, why and for how much. The website contains publications, an Impact Evaluations database, an expert roster and other useful resources.

International Organisation for Cooperation in Evaluation (IOCE)

<http://www.internationalevaluation.com/>

Created in 2003, IOCE is a loose alliance of regional and national evaluation organisations (associations, societies and networks) from around the world that collaborate to build evaluation leadership and capacity in developing countries, foster the cross-fertilisation of evaluation theory and practice around the world, address international challenges in evaluation, and assist the evaluation profession to take a more global approach to contributing to the identification and solution of world problems. Its website proposes some documents, reports, case-studies and discussion forums.

Network of Networks on Impact Evaluation (NONIE)

<http://www.worldbank.org/ieg/nonie/>

Created in 2007, NONIE is comprised of the Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC) Evaluation Network, the United Nations Evaluation Group (UNEG), the Evaluation Cooperation Group (ECG), and the International Organization for Cooperation in Evaluation (IOCE)-a network drawn from the regional evaluation associations. NONIE was formed to promote quality impact evaluation. It does not attempt to address wider monitoring and evaluation issues. Its website contains an impact evaluation database, working papers, and other resources.

Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC) Evaluation Network

www.oecd.org/dac/evaluationnetwork

This Network is a subsidiary body of the Development Assistance Committee (DAC). Its purpose is to increase the effectiveness of international development programmes by supporting robust, informed and independent evaluation. The Network brings together evaluation managers and specialists from OECD development cooperation agencies and multilateral development institutions (regional and international development banks, UNDP). Its website contains documents and publications, a calendar of evaluation events and links to other evaluation websites.

United Nations Evaluation Group (UNEG)

<http://www.uneval.org/>

UNEG brings together the units responsible for evaluation in the UN agencies, funds, programmes and affiliated organisations. Its aims are to strengthen the objectivity, effectiveness and visibility of the evaluation function across the UN system and to advocate the importance of evaluation for learning, decision making and accountability. UNEG provides a forum for members to share experiences and information, discuss the latest evaluation issues and promote simplification and harmonisation of reporting practices. Its website contains the UN evaluation standards, publications and resources and a country-level evaluation database.

9.2 Regional Evaluation Associations

African Evaluation Association (AfrEA)

<http://www.afrea.org/>

AfrEA was created in 1999, as an umbrella association for the African national evaluation associations and as resource for individual evaluators in countries where associations do not yet exist. Its goal is to promote and strengthen evaluation in Africa. Its websites proposes among others monitoring and evaluation resources (by sector), discussion forums, links to national evaluation associations, and a directory of African evaluators.

American Evaluation Association (AEA)

<http://www.eval.org/>

AEA is an international professional association of evaluators devoted to the application and exploration of programme evaluation, personnel evaluation, technology, and many other forms of evaluation. It has approximately 5500 members representing all 50 states in the US as well as over 60 foreign countries. Its mission is to promote and improve evaluations. Its websites contains contacts (by topics or location), bibliographies, links and other others.

Australasian Evaluation Society (AES)

<http://www.aes.asn.au/>

AES is the primary professional organisation in Australasia for people involved in evaluation. It has over 1000 members involved in all aspects of evaluation and performance measurement. Its website contains articles, documents, stories and evaluation contacts.

European Evaluation Society (EES)

<http://www.europeanevaluation.org/>

EES was created in 1994 to promote theory, practice and utilization of high quality evaluation especially, but not exclusively, within the European countries. Its website contains useful links, news on the evaluation sector, as well as services for its members.

International Program Evaluation Network (IPEN)

<http://www.eval-net.org/>

IPEN is the regional evaluation network covering the countries of the Commonwealth of Independent States (CIS). Its website is in Russian.

Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLAC)

<http://www.relacweb.org/>

Created in 2003, ReLAC is a network of networks aiming at promoting M&E and professionalising the sector through capacity building in Latin America and the Caribbean. Its website (in Spanish) contains a few documents and links, but mostly contacts in all ReLAC Members (Latin American and Caribbean Evaluation Associations).

9.3 Development Banks Evaluation Departments

All regional and global Development Banks, as well as Financial Organisations have their own Evaluation unit or department, which are listed below. Their websites include evaluation of interventions funded by these organisations, but also more general evaluation guidelines and resources. The heads of these evaluation departments also created an evaluation group common to all these multilateral organisations (the Evaluation Cooperation Group, see below).

African Development Bank, Evaluation Reports

<http://www.afdb.org/en/documents/evaluation-reports/>

Asian Development Bank, Independent Evaluation Department (IED)

<http://www.adb.org/IED/>

European Bank for Reconstruction and Development, Evaluation Department

<http://www.ebrd.com/projects/eval/>

European Investment Bank, Operations Evaluation Department

<http://www.eib.org/projects/evaluation/index.htm>

Evaluation Cooperation Group (ECG)

www.ecgnet.org

Inter-American Development Bank, Office of Evaluation and Oversight (OVE)

<http://www.iadb.org/ove/>

International Monetary Fund, Independent Evaluation Office (IEO)

<http://www.ieso-imf.org/>

The World Bank is particularly active in the field of impact evaluations. Its main evaluation body is the Independent Evaluation Group (IEG, see below), but it also launched other evaluation initiatives on specific themes or regions (DIME), or upon request of specific donor countries (SIEF).

World Bank, Development IMPact Evaluation Initiative (DIME)

<http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDEVIMPEVAINI/0..menuPK:3998281~pagePK:64168427~piPK:64168435~theSitePK:3998212,00.html>

World Bank, Independent Evaluation Group (IEG)

<http://www.worldbank.org/ieg/>

World Bank, Spanish Impact Evaluation Fund (SIEF)

<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0..contentMDK:21419502~menuPK:384336~pagePK:148956~piPK:216618~theSitePK:384329,00.html>

9.4 Research organisations on monitoring, evaluation, assessment and development

Centre for Global Development, Evaluation Gap Working Group

http://www.cgdev.org/section/initiatives/_active/evalgap

The Evaluation Gap Working Group was convened by the Global Health Policy Research Network as an initiative of the Centre for Global Development, to address the problem of the lack of knowledge about the effectiveness of social programmes in low- and middle-income countries. The report 'When Will We Ever Learn? Improving Lives through Impact Evaluation' can be downloaded from their webpage.

Evaluation Unit, International Development Research Centre (IDRC)

http://www.idrc.ca/en/ev-26266-201-1-DO_TOPIC.html

IDRC was created by the Parliament of Canada in 1970 to help developing countries use science and technology to find practical, long-term solutions to their social, economic, and environmental problems. Its Evaluation Unit works to strengthen the use, influence and quality of evaluation by engaging in strategic evaluations, evaluation capacity

development, evaluation tools and methods development and use, organizational learning processes. IDRC participated to the development of the outcome mapping method.

International NGO training and research Centre (INTRAC)

<http://www.intrac.org/>

Created in 1991, INTRAC is a non-profit organisation supporting NGOs and civil society organisations working in the international development and relief sector, by helping to explore policy issues, and by strengthening management and organisational effectiveness. Their website contains practical publications on monitoring and evaluation. INTRAC also proposes M&E trainings.

International Program for Development Education Training (IPDET)

<http://www.ipdet.org/>

The IPDET is an executive training programme in development evaluation. IPDET is a collaboration of the Independent Evaluation Group of the World Bank and Carleton University (Canada). The websites gives all the trainings information.

Jameel Abdul Latif Poverty Action Lab (J-PAL)

<http://www.povertyactionlab.com/>

J-PAL is a research organisation affiliated with the Massachusetts Institute of Technology, focusing on the impact evaluation of development projects through randomized controlled trials. Their website contains news, contacts, descriptions of their projects, downloadable data and publications.

Overseas Development Institute (ODI)

<http://www.odi.org.uk/>

ODI is Britain's leading independent think tank on international development and humanitarian issues. Their mission is to inspire and inform policy and practice leading to the reduction of poverty, the alleviation of suffering and the achievement of sustainable livelihoods in developing countries. ODI's evaluation experience includes the promotion of qualitative methods. Their website contains information on their activities, publications and resources classified by themes (including water and sanitation).

9.5 Other Monitoring, evaluation and assessment websites

CGIAR Group on Institutional Learning And Change (ILAC)

<http://www.cgiar-ilac.org/>

The ILAC Initiative brings together a group of national and international partners involved in collaborative applied R&D programs related to pro-poor agricultural innovation. Although not WASH-specific, this website contains resources (case-studies, tools and methods, links, forum, library of publications, evaluations, videos, etc) on applied research and evaluation that could be applied to the WASH sector in the future.

EvaluationWiki, Evaluation Resource Institute (ERI)

<http://www.evaluationwiki.org/>

EvaluationWiki was founded in 2006 by the non-profit organization Evaluation Resource Institute (ERI). The mission of EvaluationWiki is to make freely available a compendium of up-to-date information and resources to everyone involved in or interested in the science and practice of evaluation. It provides among others general resources on evaluation theory, practice and procedures that were written in a participatory manner. Registered users can edit articles, contribute to the bibliography, post comments or other information.

Monitoring and Evaluation News (Mande)

<http://mande.co.uk/>

Mande is a news service focusing on developments in monitoring and evaluation methods relevant to development programmes with social development objectives. It is managed by Rick Davies, the inventor of various qualitative monitoring and evaluation methods such as MSC. This website describes various M&E methods, proposes news and email forums on M&E subjects but also enables visitors to post their own messages, news and comments.

Outcome mapping Learning Community

<http://www.outcomemapping.ca/>

The Outcome Mapping Learning Community is an informal group of over a thousand members from around the world. It acts largely as a dynamic platform for sharing knowledge and experiences relating to Outcome Mapping. Members come together to solve problems, to showcase and trade their discoveries and good practices, and to support one another in applying OM. The website has a participatory library on Outcome Mapping with publications and contributions classified by theme and types of documents, discussion forums and mailing lists.

10. TOP contacts

Centre for Global Development (CDG)

The Evaluation Gap Working Group

Dr. Ruth Levine, Vice President Programs & Operations, Senior Fellow: rlevine@cgdev.org

1776 Massachusetts Ave. NW

Third Floor

Washington DC 20036

Tel: +1 (0) 202 416-0700

<http://www.cgdev.org/>

Consultative Group on International Agricultural Research (CGIAR) Group

Institutional Learning and Change (ILAC) Initiative

c/o Bioversity International

Via dei Tre Denari 472a

00057 Maccarese (Fiumicino)

Rome, Italy

Tel: (39) 066118253

Email: ilac@cgiar.org

www.cgiar-ilac.org

Newsbrief: <http://www.cgiar-ilac.org/content/ilac-brief>

The Evaluation Cooperation Group (ECG)

ECG Secretariat

c/o Director General

Operations Evaluation Department

Asian Development Bank

6 ADB Avenue, Mandaluyong City

1550 Metro Manila

Philippines

Tel: +63 2 6324100

https://wpgp1.adb.org/QuickPlace/ecg/Main.nsf/h_Toc/73ffb29010478ff348257290000f43a6/?OpenDocument

International Development Research Centre (IDRC)

Dr. Fred Carden, Director, Evaluation

Ms. Sarah Earl, Senior Program Officer

150 Kent Street

PO Box 8500

Ottawa, Ontario

Canada

K1G 3H9

Tel: +1(0)613 236 6163

evaluation@idrc.ca

http://www.idrc.ca/directory/office_info.php?ID=9

Institute for Development Studies (IDS)

Dr. Robert Chambers: R.Chambers@ids.ac.uk

Institute of Development Studies,

at the University of Sussex,

Brighton,

BN1 9RE

UK

Tel: +44 (0)1273 606261

<http://www.ids.ac.uk/go/home>

International Initiative for Impact Evaluation (3ie)

Mr. Howard White, Executive Director: hwhite@3ieimpact.org

Global Development Network

Post Box No. 7510

Vasant Kunj P.O.

New Delhi – 110070, India

Website: <http://www.3ieimpact.org/>

International Organisation for Cooperation in Evaluation (IOCE)

Mrs. Oumoul Ba Tall, President: oktconsult@yahoo.fr

Mr. Burt Perrin, Vice President: Burt_Perrin@Compuserve.com

Phone: +1 (0)613 722 8796

<http://www.internationalevaluation.com/>

Regional or International evaluation organisations:

http://www.internationalevaluation.com/members/reg_intl_organizations.shtml

National evaluation organisations:

http://www.internationalevaluation.com/members/national_organizations.shtml

IRC International Water and Sanitation Centre

Dr. Kristof Bostoen: bostoen@irc.nl

P.O. Box 82327

2508 EH The Hague, The Netherlands

Tel: + 31 (0)70 3044000

Website: <http://www.irc.nl>

International Water Management Institute (IWMI)

Mrs. Nadia Manning-Thomas: n.manning@CGIAR.ORG

c/o ILRI Campus

Bole Sub-City, P.O. Box 5689

Addis Ababa, Ethiopia
Tel: + 251 (0)11 6463251 (ext. 2155)
Websites: www.iwmi.org and www.ks-cgiar.org

Learning by Design

Dr. Irene Guijt: irene.guijt@wur.nl
Bredeweg 31
6668 AR Randwijk
The Netherlands

J-PAL, Jameel Abdul Latif Poverty Action Lab

Ms. Esther Duflo: eduflo@mit.edu
Massachusetts Institute of Technology
30 Memorial Drive
Building E60, Room 275
Cambridge, Massachusetts, 02142
Tel: +1 617-258-7013
Website: <http://www.povertyactionlab.com/>

London School Tropical Medicine & Hygiene

Prof. Sandy Cairncross: sandy.cairncross@lshtm.ac.uk
London School of Hygiene & Tropical Medicine
Keppel Street
London WC1E 7HT
United Kingdom
Tel: +44 (0) 20 7636 8636
<http://www.lshtm.ac.uk/>

MandE, Monitoring and Evaluation News

Mr. Rick Davies: rick@mande.co.uk
Tel: +44 (0)1223 841367
Website: <http://mande.co.uk/>

Network of Networks on Impact Evaluation (NONIE)

Secretariat c/o Victoria Gunnarsson
The World Bank Group
Mailstop MSC 9-005
700 19th Street., NW
Washington, DC 20431
USA
<http://www.worldbank.org/ieg/nonie/>

Organisation for Economic Co-operation and Development's

Development Assistance Committee (OECD/DAC)

Network on Development Evaluation

2, rue André Pascal

75775 Paris Cedex 16

France

Tel.: +33 (0)1 45 24 82 00

http://www.oecd.org/department/0,3355,en_2649_34435_1_1_1_1_1,00.html

Outcome Mapping Online Community

info@outcomemapping.ca

<http://www.outcomemapping.ca/about/index.php>

Overseas Development Institute (ODI)

Research and Policy in Development (RAPID) programme

Mr. Simon Hearn: s.hearn@odi.org.uk

Mr. Harry Jones: h.jones@odi.org.uk

111 Westminster Bridge Road

London SE1 7JD

United Kingdom

Tel: +44 (0)20 7922 0300

<http://www.odi.org.uk/default.asp>

Participatory Impact Pathways Analysis (PIPA) Wiki

Mr. Boru Douthwaite

b.douthwaite@cgiar.org

RMIT, Royal Melbourne Institute of Technology

Prof. Patricia J. Rogers: patricia.rogers@rmit.edu.au

GPO Box 2476V

Melbourne VIC 3001, Australia

Tel: +61 3 9925 2854

Website: <http://www.rmit.edu.au/>

The United Nations Evaluation Group (UNEG)

UNEG Secretariat

One United Nations Plaza

DC1-480

New York, NY 10017

USA

<http://www.uneval.org/>

11. References

Abd-Alhadi, F. T. et al. (2006). 'The Effect of Process Documentation on Building the Capacities of EMPOWERS Stakeholders for Local Water Governance'. Paper presented at the Symposium on Sustainable Water Supply and Sanitation: Strengthening Capacity for Local Governance, 26-28 September 2006, Delft, the Netherlands. Retrieved 26 February 2009 from the World Wide Web:

<http://www.irc.nl/page/31016>

Almedom, A. M.; Blumenthal, U. J. and Manderson, L. (1997). *Hygiene Evaluation Procedures : Approaches and Methods for Assessing Water- and Sanitation-related Hygiene Practice*. Boston, USA, International Nutrition Foundation for Developing Countries. Retrieved 19 March 2009 from the World Wide Web:

<http://nzdl.sadl.uleth.ca/cgi-bin/library?e=d-00000-00---off-0ccgi--00-0--0-10-0---0---0prompt-10---4-----0-1l--11-en-50---20-about---00-0-1-00-0-0-11-1-0utfZz-8-00&cl=CL3.3&d=HASH010feb15b38138ec71449a7c&x=1>

Almedon, A. (2003). 'Participatory hygiene evaluation : a means to an end, or an end in itself?' In: *Waterlines*, vol. 22, no. 1, p. 2-4.

Baker, J. L. (2000). *Evaluating the Impact of Development Projects on Poverty : A Handbook for Practitioners*. Washington, D. C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

<http://siteresources.worldbank.org/INTISPMA/Resources/handbook.pdf>

Baltazar, J. C.; Nadera, D. P. and Victora, C. G. (2002). 'Evaluation of the National Control of Diarrhoeal Disease Programme in the Philippines, 1980–93'. In: *Bulletin of the World Health Organization*, vol. 80, no. 8, p. 637-643. Retrieved 19 March 2009 from the World Wide Web:

[http://www.who.int/bulletin/archives/80\(8\)637.pdf](http://www.who.int/bulletin/archives/80(8)637.pdf)

Bamberger, M. (2006). *Conducting quality impact evaluations under budget, time and data constraints*. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

[http://inweb90.worldbank.org/oed/oeddoelib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/\\$file/conduct_qual_impact.pdf](http://inweb90.worldbank.org/oed/oeddoelib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/$file/conduct_qual_impact.pdf)

Barreto, M. L. et al. (2007). 'Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil : assessment by two cohort studies'. In: *Lancet*, vol. 370, no. 9599, p. 1622–1628. Retrieved 19 March 2009 from the World Wide Web:

<http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2212752&blobtype=pdf>

Ba Tall, O. K. (2008). *IOCE Participation to the MES-IDEAS Workshop*. Midrand, International Development Evaluation Association. Retrieved 19 March 2009 from the World Wide Web:

http://www.ideas-int.org/Documents/Oumou_Ba_Tall.doc

Belli, P. et al. (1997). *Handbook on economic analysis of investment operations*. Washington, D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

<http://www.preventionweb.net/english/professional/publications/v.php?id=1064>

Bostoen, K. (2007). *Measuring Access and Practice : Designing a Survey Methodology for the Hygiene, Sanitation and Water Sector*. Infectious and Tropical Diseases. University of London, London School of Hygiene and Tropical Medicine. PhD: 637.

Bostoen, K. (2008). *Dichotomy between global and local monitoring*. London, UK, London School of Hygiene and tropical Medicine. Unpublished.

Blum, D. and Feachem, R. G. (1983). 'Measuring the Impact of Water Supply and Sanitation Investments on Diarrhoeal Diseases : Problems of Methodology.' In: *International Journal of Epidemiology*, vol. 12, no. 3, p. 357-65.

Blum, D. et al. (1990). 'The Imo State (Nigeria) Drinking Water Supply and Sanitation Project, 1. Description of the project, evaluation methods, and impact on intervening variables'. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 84, issue 2, p. 309-315. Retrieved 19 March 2009 from the World Wide Web:

<http://www.irc.nl/docsearch/title/115763>

Briscoe, J.; Feachem, R. G. and Rahman, M. (1985). *Measuring the impact of water supply and sanitation facilities on diarrhoea morbidity ; prospects for case-control methods*. Geneva, Switzerland, World Health Organisation, Environmental Health Division.

Cabinet Office (2003). *The Magenta Book. Guidance Notes for Policy Evaluation and Analysis. Chapter 1: What is Policy Evaluation?* London, UK, Government Chief Social Researcher's Office, Prime Minister's Strategy Unit, Cabinet Office. Retrieved 19 March 2009 from the World Wide Web:

http://www.nationalschool.gov.uk/policyhub/downloads/Chapter_1.pdf

Cairncross, S. (1990). 'Health impacts in developing countries : new evidence and new prospects'. In: *Journal of the Institution of Water and Environmental Management*, vol. 4, no. 6, p. 571-577.

Cairncross S. and Kochar V. J. (1994). *Studying Hygiene Behaviour : Methods, Issues and Experiences*. London, UK, Sage Publications.

Cairncross, S. and Kolsky, P. J. (1997). 'Re: Water, Waste, and Well-being : A Multicountry Study'. In: *American Journal of Epidemiology*, vol. 146, no. 4, p 359-360. Retrieved 19 March 2009 from the World Wide Web:

<http://aje.oxfordjournals.org/cgi/reprint/146/4/359>

Cairncross, S. (1999). *Measuring the health impact of water and sanitation*. Well Factsheet. Loughborough, UK, Water Engineering Development Centre. Retrieved 19 March 2009 from the World Wide Web:

<http://www.lboro.ac.uk/well/resources/fact-sheets/fact-sheets-htm/mthiws.htm>

Casella, D.; James, A. J. and Jayaweera, P. (Undated). *Qualitative Information Systems, Making sure every voice counts*. The Hague, the Netherlands, International Water and Sanitation Centre. Retrieved 24 February 2009 from the World Wide Web:

<http://www.irc.nl/page/37610>

Catley, A.; Burns, J.; Abebe, D. and Suji, O. (2008). *Participatory Impact Assessment : a Guide for Practitioners*. Medford, Feinstein International Center, Tufts University. Retrieved 19 March 2009 from the World Wide Web:

<https://wikis.uit.tufts.edu/confluence/display/FIC/Participatory+Impact+Assessment--+a+Guide+for+Practitioners>

Chiller, T. M. et al. (2006). 'Reducing diarrhoea in Guatemalan Children : randomized controlled trial of flocculant-disinfectant for drinking-water'. In: *Bulletin of the World Health Organization*, vol. 84, no. 1, p. 28-35. Retrieved 19 March 2009 from the World Wide Web:

<http://www.who.int/bulletin/volumes/84/1/28.pdf>

Clasen, T. et al. (2007). 'Interventions to improve water quality for preventing diarrhoea : systematic review and meta-analysis'. In: *British Medical Journal* online. Retrieved 19 March 2009 from the World Wide Web:

<http://www.bmj.com/cgi/content/full/334/7597/782?view=long&pmid=17353208>

Colford, J.M. Jr. et al. (2005). 'A Randomized, Controlled Trial of In-Home Drinking Water Intervention to Reduce Gastrointestinal Illness'. In: *American Journal of Epidemiology*, vol. 161, no. 5, p. 472-482. Retrieved 19 March 2009 from the World Wide Web:

<http://aje.oxfordjournals.org/cgi/reprint/161/5/472>

DAC (2002). *Glossary of Key Terms in Evaluation and Results based Management*. Paris, France, Organisation for Economic Cooperation and Development. Retrieved 19 March 2009 from the World Wide Web:

<http://www.oecd.org/dataoecd/29/21/2754804.pdf>

Davies, R. and Dart, J. (2005). *The Most Significant Change (MSC) Technique*. Cambridge, UK, Monitoring and Evaluation News. Retrieved 19 March 2009 from the World Wide Web:

<http://www.mande.co.uk/docs/MSCGuide.pdf>

Davies, R. (2008). *Network models and Social Frameworks for representing project contexts, plans and outcomes*. Cambridge, UK, Monitoring and Evaluation News.

Retrieved 19 March 2009 from the World Wide Web:

<http://mande.co.uk/special-issues/network-models/>

Davies, R. (2009). *Rick's methods*. This blog presents various methods to monitor, evaluate and assess change in development projects. Retrieved 19 March 2009 from the World Wide Web:

<http://mande.co.uk/special-issues/>

Douthwaite, B.; Alvarez, S.; Thiele, G. and Mackay, R. (2008). *Participatory Impact Pathways Analysis : A practical method for project planning and evaluation*. Institutional Learning and Change Initiative, Brief n° 17. Retrieved 11 May 2009 from the World Wide Web:

http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief17_PIPA.pdf

Duflo, E. and Kremer, M. (2003). *Use of Randomization in the Evaluation of Development Effectiveness*. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

<http://econ-www.mit.edu/files/765>

Earl, S.; Carden, F. and Smutylo, T. (2001a). *Outcome Mapping, Building Learning and Reflection into Development Programs*. Ottawa, Canada, International Development Research Centre. Retrieved 19 March 2009 from the World Wide Web:

<http://www.idrc.ca/openebooks/959-3/>

Earl, S.; Carden, F. and Smutylo, T. (2001b). *Outcome Mapping, the challenges of assessing development impacts*. Ottawa, Canada, International Development Research Centre. Retrieved 19 March 2009 from the World Wide Web:

http://www.idrc.ca/uploads/user-S/10945665201om_pamplet_final.ppt

EES (2007). *The importance of methodologically diverse approach to impact evaluation : specifically with respect to development aid and development interventions*. Nijkerk, the Netherlands, European Evaluation Society Secretariat. Retrieved 19 March 2009 from the World Wide Web:

<http://www.europeanevaluation.org/download/?noGzip=1&id=1969403>

Esrey, S. A. (1996). 'Water, Waste, and Well-Being: A Multicountry Study'. In: *American Journal of Epidemiology*, vol. 143, no. 6, p. 608-623. Retrieved 19 March 2009 from the World Wide Web:

<http://aje.oxfordjournals.org/cgi/reprint/143/6/608>

Estis, A. (1998). 'The balanced Scorecard, applying a private sector technique to the public sector'. Paper presented at the 1998 Conference of the Association for Public Policy Analysis and Management. Retrieved 19 March 2009 from the World Wide Web:
<http://www.scribd.com/doc/7137694/BSC2200-Balanced-Scorecard-in-Public-Sector>

Fewtrell, L. and Colford, J. M. Jr. (2004). *Water, sanitation and hygiene : interventions and diarrhoea. A systematic review and meta-analysis*. Washington D.C., USA, The International Bank for Reconstruction and Development, the World Bank. Retrieved 19 March 2009 from the World Wide Web:
<http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1095698140167/Fewtrell&ColfordJuly2004.pdf>

Galiani, S.; Gertler, P. and Schargrodsy, E. (2002). *Water for Life : The Impact of the Privatization of Water Services on Child Mortality*. Berkeley, USA, University of California. Retrieved 19 March 2009 from the World Wide Web:
http://faculty.haas.berkeley.edu/gertler/working_papers/Water%20for%20Life%20June30.pdf

Gill, M. and Spriggs, A. (Undated). *The Development of Realistic Evaluation Theory through the Evaluation of National Crime Prevention Programmes*. Ottawa, Canada, Canadian Evaluation Society. Retrieved 19 March 2009 from the World Wide Web:
http://www.evaluationcanada.ca/distribution/20021010_gill_martin_spriggs_angela.pdf

Habicht, J. P.; Victora, C. G. and Vaughan, J. P. (1999). 'Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact'. In: *International Journal of Epidemiology*, vol. 28, no. 1, p. 10-18. Retrieved 19 March 2009 from the World Wide Web:
<http://ije.oxfordjournals.org/cgi/reprint/28/1/10>

Hearn, S. (2008). *Outcome mapping, complexity and aid*. London, US, Overseas Development Institute. Retrieved 19 March 2009 from the World Wide Web:
www.odi.org.uk/RAPID/events/Complexity/docs/Method_OM.ppt

Hunter, P. R.; Waite, M. and Ronchi, E. (eds.) (2002). *Drinking water and infectious disease : establishing the link*. Retrieved 19 March 2009 from the World Wide Web:
http://books.google.fr/books?id=79yVekFYTCkC&pg=PA62&lpg=PA62&dq=hunter+%22case-control%22&source=bl&ots=gl_Sn-pGgk&sig=Ah7EGd5EzaEq7tl8UAkZ9OglGF4&hl=fr&ei=OM2bSbOFDo-C-gbi-azWBA&sa=X&oi=book_result&resnum=1&ct=result#PPP1,M1

Hospes, O. (2008). 'Evaluation evolution?' In: *The Broker online*, issue 8, p. 24-26. Retrieved 19 March 2009 from the World Wide Web:
<http://www.thebrokeronline.eu/en/articles/Evaluation-evolution#f2>

Hutton, G. and Haller, L. (2004). *Evaluation of the Costs and Benefits of Water and Sanitation Improvements at the Global Level*. Geneva, Switzerland, World Health Organisation. Retrieved 19 March 2009 from the World Wide Web:
http://www.who.int/water_sanitation_health/wsh0404.pdf

IEG (2008). *Learning what works from impact evaluation : a review of the nutrition and water supply and sanitation sectors*. IEG Impact Evaluation Study. Washington D. C, USA, The World Bank. Unpublished Report.

Jalan, J. and Ravallion, M. (2001). *Does piped water reduce diarrhoea for children in rural India?* Policy Research Working Paper 2664. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:
http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2001/11/22/000094946_01091104014290/Rendered/PDF/multi0page.pdf

JPAL (2007). *Household water connections in Urban Morocco*. Cambridge, USA, Jameel Abdul Latif Poverty Action Lab. Retrieved 19 March 2009 from the World Wide Web:
<http://www.povertyactionlab.org/projects/project.php?pid=93>

Kanyesigye, J.; Anguria, J.; Niwagaba, E. and Williamson, T. (2004). *Are national water and sanitation objectives being achieved on the ground? A review of service delivery, planning monitoring & evaluation in Tororo and Wakiso districts*. London, UK, WaterAid. Retrieved 19 March 2009 from the World Wide Web:
http://www.wateraid.org/documents/plugin_documents/watersanitationobjectives.pdf

Kirchhoff, L. V. et al. (1985). 'Feasibility and efficacy of in-home water chlorination in rural north-eastern Brazil'. In: *Journal of Hygiene*, vol. 94, no. 2, p. 173-180.

Lenton, R.; Wright, A. M. and Lewis, K. (lead authors) (2005). *Health, dignity and development : what will it take?* UN Millenium Project, Task Force on Water and Sanitation. New York, USA, United Nations Development Programme. Retrieved 19 March 2009 from the World Wide Web:
<http://www.unmillenniumproject.org/documents/WaterComplete-lowres.pdf>

Luby, S. P. et al. (2005). 'Effect of handwashing on child health : a randomised controlled trial'. In: *The Lancet*, vol. 366, issue 9481, p. 225-233. Retrieved 19 March 2009 from the World Wide Web:
[http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(05\)66912-7/fulltext#article_upsell](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(05)66912-7/fulltext#article_upsell)

Mowles, C.; Stacey, R. and Griffin, D. (2008). 'What contribution can insights from the complexity sciences make to the theory and practice of development management?' In: *Journal of International Development*, vol. 20, issue 6, p. 804-820.

-
- Narayan, D. (1993). *Participatory evaluation : Tools for managing change in water and sanitation*. World Bank Technical Paper 207. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:
http://books.google.fr/books?id=SA_CG4ygWuQC&dq=narayan+managing+change&prints=ec=frontcover&source=bl&ots=kqv3iLgm9F&sig=NGLWxxhzgKgU847XIZW0jEhxKvl&hl=fr&ei=2Sq2SYH5LITR-Aaa37iICw&sa=X&oi=book_result&resnum=1&ct=result
- NEF (2009). *Proving and improving*. London, UK, New Economics Foundation. Retrieved 19 March 2009 from the World Wide Web:
<http://www.proveandimprove.org/new/provingandimproving.php>
- NONIE (2008). *Draft NONIE statement on impact evaluation*. Washington, D.C., UK, Network Of Networks on Impact Evaluation. Retrieved 19 March 2009 from the World Wide Web:
www.worldbank.org/ieg/nonie/docs/IE_statement.doc
- NZAid (2007). *Logical framework approach, developing a logical framework matrix*. Wellington, New Zealand's International Aid and Development Agency. Retrieved 19 March 2009 from the World Wide Web:
<http://nzaidtools.nzaid.govt.nz/?q=logical-framework-approach/developing-logical-framework-matrix>
- OECD (2006). *Outline of principles of impact evaluation, part I : Key concepts*. Paris, France, Organisation for Economic Cooperation and Development. Retrieved 19 March 2009 from the World Wide Web:
<http://www.oecd.org/dataoecd/46/16/37671602.pdf>
- Parker, P. E. et al.(2001). 'Differentiating assessment from evaluation as continuous improvement tools'. Paper presented at the 31st ASEE/IEEE Frontiers in Education Conference in Reno. Retrieved 19 March 2009 from the World Wide Web:
<http://www.fie-conference.org/fie2001/papers/1462.pdf>
- Plummer, J. and Slaymaker, T. (2007). *Rethinking governance in water services*. Working Paper 284. London, UK, Overseas Development Institute. Retrieved 19 March 2009 from the World Wide Web:
<http://www.odi.org.uk/resources/odi-publications/working-papers/284-rethinking-governance-water-services.pdf>
- Postma, L.; James, A. J. and Van Wijk C. (2004). 'QIS : A New Participatory Management Tool to Assess and Act on Field Reality'. Paper presented at the 30th WEDC International Conference, Vientiane, Lao PDR. Retrieved 19 March 2009 from the World Wide Web:
<http://www.irc.nl/page/37607>

Poulos, C.; Pattanayak, S.K. and Jones, K. (2006). 'A Guide to Water and Sanitation Sector Impact Evaluation'. In: *Doing Impact Evaluation*, Number 4. Washington D.C., USA, The World Bank, Poverty Reduction and Economic Management, Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation. Retrieved 19 March 2009 from the World Wide Web:

http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_04.pdf

Ramalingam, B.; Jones, H.; Reba, T. and Young, J. (2008). *Exploring the science of complexity : Ideas and implications for development and humanitarian efforts*. ODI Working Paper 285. London, UK, Overseas Development Institute. Retrieved from the World Wide Web:

<http://www.odi.org.uk/resources/odi-publications/working-papers/285-science-complexity.pdf>

Ravallion, M. (1999). *The mystery of the vanishing benefits : Ms. Speedy Analyst's introduction to evaluation*. Policy Research Working Paper 2153. Washington D.C., USA, The World Bank. Poverty Research Group, Policy and Human Resources. Retrieved 6 February 2009 from the World Wide Web:

<http://www.worldbank.org/html/dec/Publications/Workpapers/wps2000series/wps2153/wps2153.pdf>

Ravallion, M. (2008). *Introduction to the concepts and methods of Impact Evaluation*. Washington D.C., USA, Development Research Group, World Bank. Retrieved 19 March 2009 from the World Wide Web:

http://siteresources.worldbank.org/INTISPMA/Resources/Training-Events-and-Materials/449365-1208956735213/Day1_Session3&4_IEmethods_MRavallion.ppt

Roduner, D.; Schläppi, W. and Egli, W. (2008). *Logical Framework Approach and Outcome Mapping, a constructive attempt of synthesis*. Retrieved 19 March 2009 from the World Wide Web:

http://www.outcomemapping.ca/forum/files/Discussion_Paper_OM_LFA_Synthesis_2008-1_126.pdf

Rogers, P. and Hall, A. W. (2003). *Effective Water Governance*. TEC background paper no 7. Global Water Partnership Technical Committee. Stockholm, Sweden, Global Water Partnership. Retrieved 19 March 2009 from the World Wide Web:

<http://www.gwpforum.org/gwp/library/TEC%207.pdf>

Rogers, P. J. (2008). 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions'. In: *Evaluation*, vol. 14, no. 1, p. 29-48.

Sanders, I. (2003). *What is complexity?* Washington, USA, Washington Center for Complexity and Public Policy. Retrieved 19 March 2009 from the World Wide Web:

http://www.complexsys.org/pdf/what_is_complexity.pdf

Sanders, J. R. and The Joint Committee on Standards for Educational Evaluation (1994). 'Glossary of Program Evaluation Terms'. In: *The Program Evaluation Standards*, 2nd Edition. Newbury Park, Sage Publications Inc. Retrieved 19 March 2009 from the World Wide Web:

<http://ec.wmich.edu/glossary/prog-glossary.htf#Table%20of%20Contents>

Savedoff, W. D.; Levine, R. and Birdsall, N. (eds.) (2006). *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Washington D.C., USA, Center for Global Development. Retrieved 19 March 2009 from the World Wide Web:

www.cgdev.org/files/7973_file_WillWeEverLearn.pdf

Schouten, T. (2006). *Scaling up Community Management of Rural Water Supply*. Well Factsheet. Loughborough, UK, Water Engineering Development Centre. Retrieved 19 March 2009 from the World Wide Web:

<http://www.lboro.ac.uk/well/resources/fact-sheets/fact-sheets-htm/Scaling%20up.htm>

Schouten, T. (2007a). *Learning Alliance Briefing No 6 : Process Documentation*. Briefing note prepared for a SWITCH training workshop on process documentation, Lodz, Poland, 1-5 July 2007. Retrieved 19 March 2009 from the World Wide Web:

http://www.switchtraining.eu/fileadmin/template/projects/switch_training/db/event_upload_older/34/Briefing_Note_6_-_Process_documentation.pdf

Schouten, T. (2007b). *Introduction to process documentation for learning alliances and action research based on background paper SWITCH Briefing No 6*. Retrieved 19 March 2009 from the World Wide Web:

http://www.switchurbanwater.eu/outputs/pdfs/WP6-2_CLOD_PRS_Day_1_introduction_LO_Jul07.pdf

Schouten, T.; Mized, B. and Al-Zoubi, R. (2007). *The Inside Story, Process Documentation experiences from EMPOWERS*. Amman, Inter-Islamic Network on Water Resources Development and Management. Retrieved 19 March 2009 from the World Wide Web:

<http://www.project.empowers.info/page/3287>

Schultzberg, G.; Faechem, R.G. and MacCormack, C. (1983). 'One step at a time : WHO's Minimum Evaluation Procedure for water supply and sanitation projects'. In: *Waterlines*, vol. 2, no. 1, p. 19-21.

Sijbesma, C. and Postma, L. (2008). Quantification of Qualitative Data in the Water Sector : The Challenges. In: *Water International*, vol.33, no.2, p. 1-12. Routledge, Taylor & Francis Group. Retrieved 19 March 2009 from the World Wide Web:

<http://www.irc.nl/page/45816>

Slaymaker, T. and Newborne, P. (2004). *Implementation of Water Supply & Sanitation Programmes under PRSPs : Synthesis of research findings from sub-Saharan Africa*. London, UK, Overseas Development Institute. Retrieved 19 March 2009 from the World Wide Web:

<http://www.odi.org.uk/wpp/RESOURCES/project-reports/04-Watsan-PRSP.pdf>

Smits, S.; Bustamante, R. and Butterworth, J. (2005). 'Integrated water resources management at the local level: the role of local government'. Paper presented at the Conference Agua 2005, Cali, Colombia, and republished in *Urban Water Supply and Sanitation : a Management Perspective*. Hyderabad, India, Icfai University Press. Retrieved 19 March 2009 from the World Wide Web:

http://www.iclei-europe.org/fileadmin/user_upload/logowater/resources/Local_govt_and_IWRM_Smits-Bustamante-Butterworth.pdf

Snowden, D. J. and Boone, M. E. (2007). 'A Leader's Framework for Decision Making'. In: *Harvard Business Review*, November 2007. Retrieved 19 March 2009 from the World Wide Web :

<http://www.mpiweb.org/CMS/uploadedFiles/Article%20for%20Marketing%20-%20Mary%20Boone.pdf>

Straight, H. S. (2002). *The difference between assessment and evaluation*. New York, USA, Binghamton University. Retrieved 19 March 2009 from the World Wide Web:

http://assessment.binghamton.edu/documents/assessment_evaluation_straight.ppt

Tilley, N. (2000). 'Realistic evaluation : an overview'. Paper Presented at the Founding Conference of the Danish Evaluation Society, September 2000. Retrieved 19 March 2009 from the World Wide Web:

<http://www.danskevalueringsselskab.dk/pdf/Nick%20Tilley.pdf>

Water Monitoring Alliance (2005). *What is Monitoring all about*. Marseille, France, World Water Council. Retrieved 19 March 2009 from the World Wide Web:

<http://www.watermonitoringalliance.net/index.php?id=782&L=1....com%2Fbramjvb%2Fc999.txt>

Weiss, C. H. (2000). 'Theory-based Evaluations : Theories of Change for Poverty Reduction Programs'. In: *Evaluation and Poverty Reduction*. Washington D. C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

http://books.google.fr/books?id=e8kiRvx7L-AC&dq=Evaluation+and+Poverty+Feinstein+Picciotto&printsec=frontcover&source=bl&ots=jokZn52sdt&sig=v4hHdb1z95Bqc4sdzZO1xIQk2nc&hl=fr&ei=FUKTSdDOO5iL-gaH8uWJCw&sa=X&oi=book_result&resnum=1&ct=result

White, H. (2006). *Impact Evaluation : The Experience of the Independent Evaluation Group of the World Bank*. Washington D.C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

[http://inweb90.worldbank.org/oed/oeddoelib.nsf/24cc3bb1f94ae11c85256808006a0046/35bc420995bf58f8852571e00068c6bd/\\$FILE/impact_evaluation.pdf](http://inweb90.worldbank.org/oed/oeddoelib.nsf/24cc3bb1f94ae11c85256808006a0046/35bc420995bf58f8852571e00068c6bd/$FILE/impact_evaluation.pdf)

WHO (1983). *Minimum Evaluation Procedure (MEP) for water and sanitation projects*. WHO-OMS: 52. ETS/83.1; CDD/OPR/83.1. Geneva, Switzerland, World Health Organization. Retrieved 19 March 2009 from the World Wide Web:

<http://www.irc.nl/docsearch/title/122983>

WHO (1998). *PHAST step-by-step guide : a participatory approach for the control of diarrhoeal diseases*. Geneva, Switzerland, World Health Organization. Retrieved 19 March 2009 from the World Wide Web:

http://www.who.int/water_sanitation_health/hygiene/envsan/phastep/en/index.html

Wikipedia (2008). *Longitudinal study*. San Francisco, USA, Wikimedia Foundation Inc. Retrieved 19 March 2009 from the World Wide Web:

http://en.wikipedia.org/wiki/Longitudinal_study

Wikipedia (2009). *Placebo (disambiguation)*. San Francisco, USA, Wikimedia Foundation Inc. Retrieved 19 March 2009 from the World Wide Web:

[http://en.wikipedia.org/wiki/Placebo_\(disambiguation\)](http://en.wikipedia.org/wiki/Placebo_(disambiguation))

Willets, J. (2004). *Most Significant Change Pilot Project, Evaluation Report*. Sydney, Australia, Institute for Sustainable Futures, University of Technology. Retrieved 30 March 2009 from the World Wide Web:

http://www.isf.uts.edu.au/publications/Willets_2004_MSC_Eval.pdf

Willets, J. and Crawford, P. (2007). 'The most significant lessons about the Most Significant Change technique'. In: *Development in Practice*, vol. 17, no. 3, p. 397-379. Retrieved 30 March 2009 from the World Wide Web:

<http://www.oxfam.org.uk/download/?download=http%3A%2F%2Fwww%2Einformaworld%2Ecom%2Fsmpp%2Fftinterface%3Fcontent%3Da778573642%26%23x26%3Bformat%3Dpdf%26%23x26%3Bmagic%3Doxfam%7Ctnn713412875cerv17n3x0s8%7C447C4E1C0C56B9AF778F3D78FABAA160%26%23x26%3Bft%3D%2Epdf>

Wolman, A. (1976). 'Book review : Measurement of the Health Benefits of Investments in Water Supply'. In: *American Journal of Public Health*, vol. 66, no. 7, p. 695. Retrieved 19 March 2009 from the World Wide Web:

<http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1653391&blobtype=pdf>

World Bank (1976). *Measurement of the Health Benefits of Investments in Water Supply*. Report no. PUN 20. Washington D.C., USA, The World Bank.

World Bank (1996). *The World Bank Participation Sourcebook*. Washington, D. C., USA, The World Bank. Retrieved 19 March 2009 from the World Wide Web:

<http://www.worldbank.org/wbi/sourcebook/sbintro.pdf>

World Bank (2009). *Evaluations designs*. Washington D.C., USA, The World Bank.

Retrieved 19 March 2009 from the World Wide Web:

<http://go.worldbank.org/N0YMFQG000>

Young, B. and Briscoe, J. (1987). 'A case-control study of the effect of environmental sanitation on diarrhoea morbidity in Malawi'. In: *Journal of Epidemiology and Community Health*, vol. 42, no. 1, p. 83-88. Retrieved 19 March 2009 from the World Wide Web:

<http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1052686&blobtype=pdf>

Zimmermann, A. and Engler, M. (1996). *Process Monitoring (ProM) : Work Document for project staff*. Eschborn, Germany, Deutsche Gesellschaft für Technische Zusammenarbeit.

Retrieved 19 March 2009 from the World Wide Web:

<http://www2.gtz.de/dokumente/bib/96-1351.pdf>

12. About IRC

IRC facilitates the sharing, promotion and use of knowledge so that governments, professionals and organisations can better support poor men, women and children in developing countries to obtain water and sanitation services they will use and maintain. It does this by improving the information and knowledge base of the sector and by strengthening sector resource centres in the South.

As a gateway to quality information, the IRC maintains a Documentation Unit and a web site with a weekly news service, and produces publications in English, French, Spanish and Portuguese both in print and electronically. It also offers training and experience-based learning activities, advisory and evaluation services, applied research and learning projects in Asia, Africa and Latin America; and conducts advocacy activities for the sector as a whole. Topics include community management, gender and equity, institutional development, integrated water resources management, school sanitation, and hygiene promotion.

IRC staff work as facilitators in helping people make their own decisions; are equal partners with sector professionals from the South; stimulate dialogue among all parties to create trust and promote change; and create a learning environment to develop better alternatives.

IRC International Water and Sanitation Centre

P.O. Box 82327

2508 EH The Hague

The Netherlands

Tel: +31 (0)70 3044000

Fax: +31 (0)70 3044044

E-mail: general@irc.nl

Internet: <http://www.irc.nl>

