



KINGDOM OF BELGIUM
Federal Public Service
**Foreign Affairs,
Foreign Trade and
Development Cooperation**

Impact measuring, the Quest of the Grail?

Ex post Impact Evaluation of four Bilateral
Cooperation projects

Synthesis report



inputs → outputs → outcomes → impact

**Federal Public Service Foreign Affairs, Foreign Trade and
Development Cooperation**

Office of the Special Evaluator of the International Cooperation

Impact measuring, the Quest of the Grail?

Ex post Impact Evaluation of four Bilateral Cooperation Projects



Synthesis Report

October 2013

This evaluation was conducted by ADE and the CRED supported by a steering committee. The Office of the Special Evaluator has assured the conformity of this evaluation report with the terms of reference.

The opinions expressed in this document represent the authors' views and do not necessarily reflect those of the FPS Foreign Affairs, Foreign Trade and Development Cooperation.

© FPS Foreign Affairs, Foreign Trade and Development Cooperation

January 2014

This report should be quoted as follows:

Office of the Special Evaluator (2013), *Impact measuring, the Quest of the Grail? – Ex Post Impact evaluation of four Bilateral Cooperation Projects – Synthesis report*, FPS Foreign Affairs, Foreign Trade and Development Cooperation, Brussels.

Graphic design: FPS Communication Service

Printing: FPS Printing Office

Evaluation no. S4/2011/04

Legal deposit: 0218/2014/003

This document is available in PDF format also in French and Dutch on the CD-ROM joined to the synthesis of the report, at the Special Evaluation Office and on the website www.diplomatie.belgium.be/en/policy/development_cooperation/evaluation

We would like to thank all those people and organisations who have contributed to this study. We would especially like to thank the steering committee members and the interviewees in the 4 ex post impact evaluations.

Table of Contents

TABLE OF CONTENTS	5
LIST OF FIGURES AND TABLES	7
ACRONYMS AND ABBREVIATIONS	9
EXECUTIVE SUMMARY	11
1. INTRODUCTION	19
2. METHODOLOGY	21
2.1 Impact, a definition to be clarified	21
2.2 One evaluation process, two methods	23
2.3 Quantitative methods: objectives and challenges.....	25
2.4 The evaluation’s practical organisation	36
3. MAIN CONCLUSIONS ON THE ACHIEVEMENT OF THE OUTCOMES AND IMPACT OF THE 3 PROJECTS	39
3.1 Evaluation of a Project supporting Technical and Vocational Education in the DRC (AETP2)	39
3.2 Evaluation of a Project for the Improvement and Reinforcement of Water Points in Senegal (PARPEBA).....	40
3.3 Impact evaluation of a Project for Small and Medium Hydraulics in Morocco (PMH)	41
4. LESSONS FROM THE USE OF MIXED METHODS FOR EVALUATING OUTCOMES AND IMPACT .	47
4.1 Lessons on the evaluation process	47
4.1.1 A quantitative evaluation system enhances qualitative analysis	49
4.1.2 A sound qualitative analysis is necessary for the rigorous implementation of quantitative methods.....	51
4.1.3 Qualitative analysis is the key to understanding a project's effects (outcomes/impact) and assessing the impact	54
4.1.4 A system that allows all actors to be considered	54
4.1.5 The feasibility and relevance of an ex post quantitative evaluation.....	54
4.2 Lessons on the projects' design, implementation and monitoring-evaluation .	62
4.2.1 Shortcomings in the interventions' design/formulation which reduces their potential impact.....	62

4.2.2	Implementation that favours disbursement and physical achievements.....	64
4.2.3	Inadequate M&E system	64
5.	RECOMMENDATIONS	67
5.1	Define a global evaluation strategy where impact evaluation has a place.....	67
5.1.1	Clarify the definitions of impact, outcome and impact evaluation	68
5.1.2	Raise awareness about impact evaluation	69
5.1.3	Plan for an annual impact evaluation programme ensuring feasibility and added value.....	69
5.1.4	Guarantee the quality of evaluations.....	71
5.2	Conduct high-quality rigorous impact evaluations	72
5.2.1	Design the evaluation and project at the same time.....	72
5.2.2	Successive steps and important elements in conducting a rigorous ex post impact evaluation	73
5.3	"Management towards outcomes and impact" for more realistic projects targeting the achievement of tangible effects.....	76
5.3.1	Incorporate impact evaluation at the time of the design and strengthen the project validation process so that unrealistic projects are not started	77
5.3.2	Practice genuine results based management (outputs but also outcomes and impact).....	77
6.	BIBLIOGRAPHY.....	79
	APPENDICES	81
APPENDIX 1:	EX POST IMPACT EVALUATION REPORT ON THE AETP2 PROJECT IN THE DRC	
APPENDIX 2:	EX POST IMPACT EVALUATION REPORT ON PARPEBA IN SENEGAL	
APPENDIX 3:	EX POST IMPACT EVALUATION REPORT ON THE PMH PROJECT IN MOROCCO	

List of Figures and Tables

LIST OF FIGURES

Figure 1: Two concepts of impact evaluation	22
Figure 2: Different definitions of the concepts in the logical framework	22
Figure 3: Diagram showing the double difference technique	34
Figure 4: Key steps in the evaluation process by mixed methods	36

LIST OF TABLES

Table 1: Summary presentation of the 3 evaluation reports.....	43
Table 2: Conditions for conducting quasi-experimental, ex ante or ex post design quantitative evaluations.	60

Acronyms and abbreviations

3ie	International initiative for impact evaluation
ADE	Analysis for Economic Decisions
AETP2	Project supporting technical and vocational education in the DRC
AfD	French Development Agency
ASUFOR	Association des usagers des forages / Association of borehole users
AUEA	Association des usagers des eaux agricoles / Association of agricultural water users
BTC	Belgian Technical Cooperation
CRED	Centre for research in economic development
DCD	Development Cooperation Directorate (OECD)
DGD	Directorate General for Development Cooperation and Humanitarian Aid
DRC	Democratic Republic of Congo
EUDN	European Development Network
IEG	Independent Evaluation Group (World Bank)
IGI	Income Generating Initiative
J-Pal	Abdoul Latif Jamel Poverty action lab
LF	Logical Framework
LRSP	Projet de laboratoire de référence en santé publique / National public health reference laboratory project
MDG	Millennium Development Goal
ME	Monitoring and evaluation
MIT	Massachusetts Institute of Technology
MM	Mixed Methods
OECD	Organisation for Economic Cooperation and Development
OSE	Office of the Special Evaluator of International Cooperation
PARPEBA	Project for the improvement and reinforcement of water points in the Arachidier Basin in Senegal
PMH	Project for small and medium hydraulics in the Tiznit region in Morocco
RCT	Randomized Control Trial
TVE	Technical and Vocational Education

Executive Summary

Introduction

The ex post impact evaluation of four projects of governmental cooperation was commissioned by the Office of the Special Evaluator (OSE) of the Belgian Federal Public Service Foreign Affairs, Foreign Trade and Development Cooperation. It was carried out by experts from ADE, a consultancy firm specialising in evaluation, and researchers from the Centre for research in economic development (CRED of the University of Namur) specialising in the application of quantitative evaluation methods. The evaluation focused on four projects in four different countries and sectors.

Strictly speaking, these were the first impact evaluations commissioned by the OSE combining qualitative and quantitative methods to assess the outcomes and impact of projects.

The objective of this evaluation is both summative and formative. On the one hand, this report summarises the lessons learned on the achievement of the outcomes and impact of the evaluated projects. On the other hand, it draws conclusions about the effectiveness of the methods used. It also promotes a discussion about the role of impact evaluation in the Belgian bilateral cooperation and suggests some methodological approaches for conducting them more rigorously.

A need to clarify the different concepts

One of the evaluation's first challenges was to revisit the concept of impact. According to the OECD, impact is characterised by its temporal dimension and is defined as the "positive and negative, long-term effects produced by a development intervention, directly or indirectly, intended or unintended". In the approach based on the intervention logic, impact is defined as being the final element in a causal sequence starting with the activities and resulting in the impact. The community conducting impact evaluation, such as the World Bank and 3ie (International Initiative for Impact Evaluation), defines impact as the effects on beneficiaries that can be attributed to the intervention. This last definition of impact refers explicitly to the outcomes as presented in a project's logical framework.

In the report, **impact** is defined as being **all the effects generated by the project at a general level** (without any reference to temporality); while **outcomes** are defined as the **effects on direct beneficiaries**.

This distinction between impact and outcomes is important, particularly since the evaluation combines qualitative and quantitative methods. Indeed, quantitative methods measure the effects on beneficiaries (outcomes) while attempting to demonstrate a causal relationship between the project and the measured effects. To be able to quantify and demonstrate the effects of the project, quantitative methods need enough reliable data reflecting the situation prior to the intervention (baseline) as well as the situation post-intervention for relevant outcome indicators. Furthermore, this data must be collected from a sample of the population that has been subject to the project on the one hand (called the treated group or target group) and on the other hand, from a sample of non-beneficiaries that is similar to the former in a significant number of characteristics prior to the intervention, but which has not benefited from the intervention (called

the counterfactual, control group or comparison group).¹ Quantitative methods do not generally issue conclusions on the impact, defined as the effects generated by the project at a general level. They are usually limited to assessing the outcomes, for which the findings can sometimes be generalised or extrapolated in order to deduce an impact measurement.

As for the qualitative methods, these support a reasoned judgement on the achievement of the outcomes and impact of the projects. Based on a documentary analysis, interviews at different levels and field observations, these methods analyse whether the conditions for achieving the outcomes and impact were met at different stages of the intervention logic summarised in the logical framework.

A learning process

This study sets itself apart through its **pilot and iterative** aspects.

Pilot, because the four evaluations are not standard project evaluations, their key objective being to assess the outcomes and ex post impact of the project using a mixed methodology. Coming from two very different worlds and using analytical techniques specific to their discipline, the experts from the consultancy firm and those from the academic research centre have made considerable efforts to share and make their findings consistent. These discussions between the experts and with the OSE and the steering committee have led to a process of mutual learning. This has ultimately produced a high-quality evaluation report per project, as well as this synthesis report. Iterative, since the four projects were evaluated using this methodology in different countries and sectors in order to learn as much as possible on both a summative and formative level. We should note that the availability and flexibility of the OSE, which commissioned these evaluations has also contributed to the success of the process which is still incomplete and should be continued as it gives rise to substantive debates.

An important characteristic is the ex post nature of these evaluations which focus on projects that have been finished for more than four years. This allows a certain distance with regard to the projects' effects on beneficiaries and the achievement of the impact. However, the lack of attention paid to outcomes and impact by the project monitoring and evaluation (M&E) system, added some difficulties such as keeping track of certain data or questioning some people years after the intervention. This created obstacles for the team thus preventing the strict application of the quantitative methods.

Despite these difficulties, the evaluators successfully used some quantitative instruments. In two of the four projects, a quasi-experimental design for the evaluation process was implemented. It was possible to select a credible counterfactual and to reconstruct baseline data from a sample of beneficiaries and a counterfactual through secondary data and/or data from respondents' memories. However, given the limits of the data collected and the sometimes questionable credibility of the counterfactual, the econometric tools could only be used to demonstrate the attribution of a minority of outcomes to the project. However, all the projects evaluated were subject to a reasoned judgement on the achievement of the outcomes that could not be handled quantitatively, as well as on the achievement of the impact.

It should be noted that the context has changed since the four projects were implemented. Efforts are currently being made (particularly by the BTC) to put greater emphasis on monitoring and evaluating outcomes and impact. However, these efforts still need to be put into practice and translated into practical measures and changes on the ground. There is no guarantee that it will now be easier to apply the quantitative method based on data collected through the M&E system.

¹ Experimental-type quantitative methods recommend the random assignment of people to a group of beneficiaries (treated group) and a group of non-beneficiaries (control group) using Randomized Control Trials (RCT).

The evaluations of three of the four projects are the subject of separate reports available in the appendix. The report on the evaluation of the public health laboratory project in Rwanda is unavailable since this evaluation could not be completed. The team did not receive all the permissions required by the Ministry of Health to collect and analyse the data. This report therefore summarises the principal conclusions in terms of achieving the outcomes and impact of the three other projects. The conclusions regarding the feasibility and added value of an evaluation of outcomes and impact are, however, taken from all the evaluation processes implemented in the four countries.

Findings on the achievement of the outcomes and impact of the 4 evaluated projects

Project supporting Technical and Vocational Education in the Democratic Republic of Congo (AETP2)

The methodology selected for this evaluation was a case study using a counterfactual in the city of Kisangani. This study was therefore based on a rigorous qualitative analysis of field observations and information collected through individual and group interviews in the beneficiary schools and the control group schools along with key people involved in this sector and/or working on the project in Belgium and in the DRC. In-depth interviews were conducted with school principals, teachers, students and graduates. Given the lack of usable quantitative data for this project and the fact that implementing a system for collecting first-hand data proved to be irrelevant further to the findings of the exploratory mission, the application of quantitative methods had to be abandoned.

The exercise carried out in schools in Kisangani shows that **the project did not have any significant effect on the quality of the education or on the number of pupils, nor did it improve the schools' capacity for self-financing or make the young graduates more employable**. This failure can in part be explained by the partial use of outputs, but particularly by the extremely unfavourable environment of this country's technical and vocational education sector. For many decades, the infrastructure has been in a deplorable condition, exacerbated by the violent conflicts that have ravaged the country. Moreover, the State does not have an adequate policy supported by suitable budgetary resources (for example, the school is mostly funded by the parents and their children). A lack of appropriate incentive systems within the institutions and the existence of adverse incentives harm the effectiveness of any action aimed at improving the quality of education in the DRC.

In reality, under the conditions highlighted, which were known when the project was designed and which, for the most part, still apply today, this project should never have taken place or should have been designed in a different way.

Project for the Improvement and Reinforcement of Water Points in the Arachidier Basin in Senegal (PARPEBA)

Due to the lack of usable baseline data, a quasi-experimental design was implemented to collect first-hand data from beneficiaries with access to tap water and households that were relatively similar to the beneficiaries but which did not have access to tap water within their village. Once a credible reference situation had been reconstructed using data collected through household surveys, it was possible to compare certain outcomes between the two groups (target group and control group) and between two periods (pre- and post-project).

The evaluation concludes that **the water supply from boreholes through taps has saved significant amounts of time and energy, especially for those beneficiaries with access to a private tap**. The time and energy saved did not apparently have an effect on the quantity of water consumed per day and per person. By contrast, these time and energy savings seem to have encouraged people not only to start up small businesses, especially during the dry season, but also to participate in local groups and associations. Furthermore, it was observed that users are strongly inclined to pay for tap water.

The new boreholes also had the effect of **improving well-being and strengthening social cohesion** (a reduction in conflicts between women, who are primarily responsible for collecting water). It would be worth studying the effect on education in greater detail because there are many exogenous factors that could explain the observed trend.

Nevertheless, **water quality in terms of both bacteria and chemicals contamination remains a real problem**. More than two thirds of the sites that were built or restored through PARPEBA have harmful fluorine and salt levels. In addition, there are many water contamination sources between the point of supply and the point of use. The population is unaware of the severity of these water quality problems. **Such an observation calls for urgent corrective measures to prevent the risk of serious health problems in rural populations in the near future.**

Project for Small and Medium Hydraulics in the Tiznit region in Morocco (PMH).

A quasi-experimental design was also used to compare a target group (farmers benefiting from the rehabilitation of the traditional irrigation system) with a control group (farmers who did not benefit from this type of rehabilitation) at two points in time (pre- and post- project) for certain outcome indicators by collecting first-hand data using household surveys.

The evaluation shows that as a result of the operations to rehabilitate the irrigation system, **the project had a positive impact on the surface area of cultivated land during periods of water stress, although no improvement was observed during flood periods**. Compared with those farmers who did not benefit from the project, the beneficiaries derived most of their income from farming activities during both dry and flood periods. This last conclusion should be treated with caution as it is based on an income assessment by the respondents themselves. The project also encouraged the spontaneous extension of the irrigation networks, especially upstream. Furthermore, it has potentially allowed the groundwater system to be recharged.

However, **the project does not seem to have encouraged either crop diversification or changes in farming practices. Therefore, it did not have the desired effect of reducing the rural exodus**. Given the random nature of floods, it would be worth monitoring and evaluating this project over several successive years with greater accuracy regarding the crop yield of a beneficiary group and an even more credible counterfactual. Indeed, despite its weaknesses, the project is of great value to the region.

Findings on the methodological approach

This study confirms the recent literature on impact evaluation: **the combination of qualitative and quantitative methods within the same evaluation process is the best possible solution for quality impact evaluations**. In the specific case of this evaluation, the process was based on the project's intervention logic and therefore validates theory-based evaluation. This approach uses quantitative methods in an attempt, within the limits of their application, to measure the outcomes and demonstrate their attribution to the project; while using qualitative methods to support a reasoned judgement on the achievement of the impact, but also on the achievement of the outcomes for which the application of quantitative methods is not technically possible. A rigorous qualitative analysis of the broader context provides an understanding of which mechanisms and factors have promoted or inhibited a project's effect. **An evaluative process that combines qualitative and quantitative tools therefore helps to identify "what worked" and "what did not work", and to understand why and how the effects were or were not achieved.**

Although it is appropriate to combine the two methods, the conditions for the optimum use of quantitative methods were not met for all four evaluated projects. **Thus, the evaluators had to adapt to the limitations imposed** by the absence of a baseline, usable monitoring data and a well-defined counterfactual. Nonetheless, although the

application of the quantitative methods was extremely limited, the exercise demonstrated that these methods enhance each other.

Firstly, notwithstanding the constraints, quantitative analysis tools were used to quantify certain outcomes and demonstrate their attribution. Secondly, the data collection system used for the quantitative methods generated a greater amount of qualitative data through the information gathered during exploratory missions, the collection of first-hand information using household surveys and a relatively long-term presence in the field. This data can be used to set out the arguments to assess the projects' effects (outcomes and impact), resulting in a more rigorous qualitative analysis. Furthermore, a good qualitative analysis is necessary for the rigorous implementation of quantitative methods. This can be used to argue for methodological choices and interpret the results found by statistical and econometric tools. Such an analysis also comments on the achievement of an impact whereas the quantitative analysis is limited to the achievement of outcomes.

The use of mixed methods also involves all the project stakeholders: the quantitative process goes into greater detail in the interviews with beneficiaries and non-beneficiaries while the qualitative approach helps to give a more accurate assessment of the roles of all the various stakeholders in an intervention.

However, there are some minimum conditions required to ensure the relevance of applying quantitative methods. In an ideal situation, through its M&E system, the project will have collected enough reliable data from the beneficiary group and the counterfactual, before and after the intervention. In all other scenarios, the feasibility of a quantitative analysis will depend on identifying a credible counterfactual, the possibility of reconstructing reliable pre-project data, partner support and the available resources. Indeed, the political aspect of a potentially negative demonstration of a project's effects should not be ignored. The rigour of the quantitative analyses may make local authorities reluctant to support the evaluation. Hence the need for partner support to implement a quantitative evaluation process.

Finally, the issue of the value for money of implementing such a data collection process and the quantitative data analyses should be addressed. If a rigorous qualitative analysis shows a low probability of observing the project's effect, then implementation of a quantitative process may be questioned since a quantitative analysis will probably not contribute anything more. In contrast, if there is a high probability of finding an effect, implementing such a system will help to quantify the effects on the beneficiaries and demonstrate the causal relationships with the intervention, something which a qualitative analysis cannot do. **More fundamentally, the cost/benefit analysis of an impact evaluation using quantitative tools will only be positive when the Belgian authorities responsible for development cooperation recognise the added value of a quantitative analysis of the outcomes.**

There are several possible options that can limit the costs of a quantitative analysis depending on the evaluation's objectives, including limiting the study to just one aspect of the project, focusing on a specific geographical area and/or making use of reliable secondary data. However, not carrying out a quantitative outcome evaluation because the implemented projects are modelled on projects that have already been evaluated as successful by other studies is very risky. Indeed, the external validity of the evaluations is often called into question. In addition, the influence of the context and intervention mechanism are two key elements which contribute to a project's success.

Findings in terms of design/formulation, implementation and monitoring-evaluation of projects

The evaluations have highlighted **shortcomings in the design, implementation and M&E system of the projects**. The evaluated projects have a weak intervention logic. The objectives are disconnected from the activities and the resources used. They are vague and poorly defined. The overall objective is often unrealistic given the intervention's local context. The weaknesses identified are concerning insofar as they

indicate that achieving the outcomes and impact was not a priority in the evaluated projects. Such shortcomings in the design also show that the logical framework has been reduced to a mere formality. Indeed, a detailed analysis of the intervention logic would undoubtedly have led to changes in the design and/or implementation of the four evaluated projects. This therefore raises the issue of the relevance in choosing certain projects.

Furthermore, the principle of alignment with national policies also raises questions when these policies are not conducive to achieving the expected outcomes and impact. More fundamentally, the ability to negotiate with the partner is called into question, along with the ability of the DGDC and the BTC to challenge any unrealistic project.

Implementation favours disbursement and physical achievements. Output-focused management is detrimental to a management approach focused on outcomes and impact. This management style undoubtedly reflects a lack of incentive to do otherwise within the Belgian bilateral aid system.

The M&E system for the four projects is insufficient and unsuitable. Even though the M&E systems for the four evaluated projects were designed appropriately, in practice they did not allow the monitoring or evaluation of the outcomes and impact. Indeed, although the outcome indicators were defined or referenced in the formulation of the M&E system, many of these indicators were never collected and where data does exist, it cannot be used quantitatively.

Recommendations

The recommendation is to **promote outcome and impact evaluations** since they can provide an adequate response to the objectives of accountability, decision-making and capitalising on experiences (sharing the lessons learned). In addition, the findings of a rigorous evaluation can be used by various stakeholders, including the donors, the partner, executing agencies and finally the beneficiaries. The decision to conduct impact evaluations must be part of the wider context of the aid cycle which takes into consideration future challenges for the Belgian development cooperation. An evaluative strategy should also take into account the cost/benefit ratio, which will continue to be low as long as evaluations do not play a more important role in Belgian development cooperation.

Promoting outcome and impact evaluations and incorporating them into a general evaluation strategy should be done while taking into account several elements. Firstly, the definitions of outcomes and impact must be clarified, at least on the level of the various stakeholders of Belgian development cooperation. Secondly, this type of evaluation must be introduced with the partners and the people responsible for the design/formulation and implementation of the projects so that it is incorporated into their project management. An annual programme to evaluate the outcomes and impact of Belgian development cooperation interventions should be planned, ensuring their added value, relevance and feasibility for each case.

The use of evaluations **that combine quantitative and qualitative methods within a single evaluation process** is recommended. To do so, sufficient time and resources are needed to recruit a team of experts in both qualitative and quantitative analysis tools in order to guarantee the quality of the evaluations. Consequently, it is important to have the resources to carry out these evaluations with the necessary rigour and the required adaptation to the realities in the field.

It is recommended that the **system for evaluating a project's outcomes and impact is developed during the project design stage**. This means reviewing the design, implementation and M&E system of the projects so that the pursuit of outcomes and impact becomes a priority in the same way as the achievement of outputs.

Designing the evaluation in the early stages of a project also allows the **logical framework to be used as a real formulation and management tool focused on the results and evaluation**. Indeed, this allows for a better consideration of the actual

causal relationships between the activities, the resources used and the objectives set, as well as a more realistic risk analysis. Among other things, this should mean that unrealistic projects are no longer validated. Designing a detailed evaluation system prior to an intervention requires the definition of relevant indicators, the specification of data collection methods before, during and after the intervention and the establishment of the types of analyses planned. Consequently, if this work is done well within the project, there can be significant savings in terms of time and money when evaluating outcomes and impact. Furthermore, the quality of the evaluations can also be significantly improved.

Finally, in order to be able to evaluate the outcomes and impact of a project using a mixed methodology, the project's implementation must be **coordinated with the evaluation system**. Sequential implementation (pipeline process) should be seriously considered so that a counterfactual can be used at less cost and more ethically, and quantitative methods can therefore be applied.

All these recommendations can only be achieved by providing the **right incentives** to every stakeholder involved in a project, in both the donor and recipient countries.

1. Introduction

The objective of this report is to summarise the lessons learned from the ex post impact evaluation of four projects of governmental cooperation commissioned by the Office of the Special Evaluator (OSE) of the Belgian Federal Public Service Foreign Affairs, Foreign Trade and Development Cooperation.

This evaluation was conducted in four different countries and sectors by experts from ADE, a consultancy firm specialising in evaluation, and researchers from the Centre for research in economic development (CRED of the University of Namur). This collaboration between a consultancy firm and an academic research centre has allowed the implementation of an original methodological approach based on the intervention logic and using both qualitative and quantitative methods. **The findings focus on the achievement of the outcomes and impact of the studied projects, the methods used and the feasibility and added value of outcome and impact evaluation.** The report also promotes a discussion on the role of impact evaluations in the Belgian bilateral aid sector and suggests some methodological approaches for conducting them.

These findings are illustrated using examples taken from the evaluation of four interventions of Belgian governmental cooperation: a drinking water access project in Senegal (PARPEBA, 2003-2008), a project on small and medium hydraulics in Morocco (PMH, 2004-2007), a project supporting technical and vocational education in the DRC (AETP2, 2005-2008) and finally a project supporting the implementation of a public health reference laboratory in Rwanda (LRSP, 2003-2008).

It should be noted that the evaluation of the project in Rwanda was not completed due to a lack of consent to collect data essential to the analysis. Indeed, the permissions required by the Ministry of Health for the collection and use of the data, and responses to the requests for such permission were not obtained. Consequently, the OSE, in agreement with the consultants, decided not to publish the report.

Strictly speaking these are the first impact evaluations carried out by the OSE. In addition, they are **ex post impact evaluations**. These evaluations aim to verify whether these projects, which have been finished for several years, achieved their objectives (in terms of outcomes and impact) and to understand the mechanisms leading to the findings of success or failure. Notwithstanding the inevitable changes that have occurred in Belgian bilateral aid since the design and implementation of these projects which ended in 2007 or 2008, for the most part, the conclusions are still relevant. They will however be modified slightly where necessary.

This report is structured as follows: Section 2 starts the debate on the definition of the term 'impact' and presents the qualitative and quantitative methods in the current impact evaluation context. Then it sets out the objectives and challenges of the quantitative methods. Lastly, this section summarises the general methodological approach applied within the framework of the four evaluations. Section 3 summarises the findings on the achievement of the outcomes and impact of three of the four evaluated projects (as already mentioned, the evaluation in Rwanda was not completed); the evaluation reports are available in separate appendices. Section 4 draws lessons from the evaluation process that combines qualitative and quantitative methods implemented to assess the projects' outcomes and impact. This process can also provide lessons about the design, implementation and monitoring of the projects. Section 5 presents the recommendations.

2. Methodology

2.1 Impact, a definition to be clarified

The lack of a clear definition of the term 'impact' is a source of misunderstanding and confusion that may hinder a serene discussion on impact evaluation and its implementation.

In the particular context of cooperation development and impact evaluation, there are three main, co-existing definitions:

- (1) The definition of impact according to the OECD refers to the temporality of an intervention's effects. In the **DAC** glossary, **impact is defined as the "positive and negative, primary and secondary, long-term effects produced by a development intervention, directly or indirectly, intended or unintended"²**. The notion of temporality leads to confusion regarding the effects to be evaluated because the time needed for the effects to appear may vary significantly depending on the type of intervention.
- (2) The second definition comes from the intervention logic as summarised in the logical framework where a development action is structured into inputs (activities), outputs (intermediate results), specific objectives and the global objective. **Impact, according to the intervention logic approach is equivalent to the global objective, or the intervention's contribution to changes at a general level.** The impact evaluation is therefore an assessment of the achievement of the global objective. It should be noted that with this approach, outcomes correspond to the specific objectives.
- (3) **Practitioners of quantitative methods use the term impact to refer to outcomes, which they define as the effects on the beneficiaries that can be attributed to the project, without concern for the temporality of these.** They therefore define impact evaluation as "the evaluation of causal effects of an intervention on outcomes irrespective of time dimension" (World Bank). The concept of attribution refers to the use of a counterfactual (see below), which leads the World Bank's independent evaluation group to suggest another definition for impact evaluation: "the impact evaluation is the counterfactual analysis of the impact of an intervention on final welfare outcomes". The term impact is used even though it clearly refers to the evaluation of an intervention's outcomes. Finally, according to H. White, "impact evaluation" is a synonym of "attribution analysis" or even "with versus without analysis" ("what happened with the program -factual record- compared to what would have happened in the absence of the program -which requires a counterfactual-", 3ie, 2013).

As illustrated in Figure 1, qualitative and quantitative methods do not pursue the same evaluation objectives. The first are interested in all the processes leading to the impact defined as the final and general effect of an intervention. The second measure a project's outcomes, defined as the effects on the direct beneficiaries and demonstrates their attribution.

² "Glossary of key terms in evaluation and results based management", OECD – DAC (2010).

Quantitative methods do not aim to issue conclusions on the effect of an intervention at a general level (impact).

Figure 1: Two concepts of impact evaluation

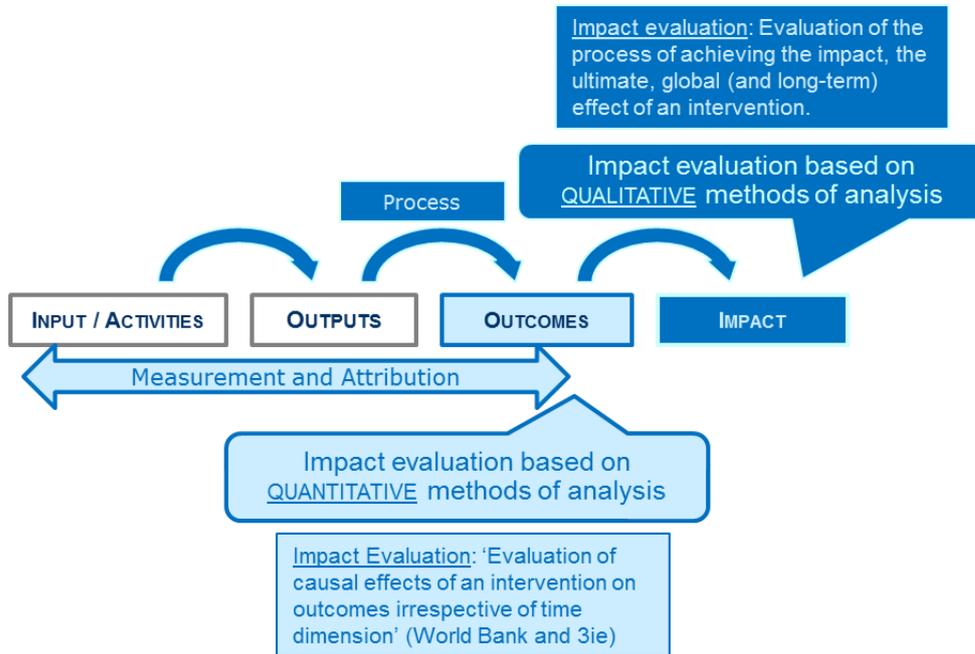


Figure 2 presents the concepts used in the logical framework with the various existing definitions. The first line of definitions corresponds to the definitions used for these evaluations. The next two lines give the DAC and intervention logic definitions respectively.

Figure 2: Different definitions of the concepts in the logical framework

	INPUTS	OUTPUTS	OUTCOMES	IMPACT
Concepts and definitions used in ADE/CRED evaluation reports	INPUTS Resources implemented and Activities undertaken with a view to producing the project's expected effects	OUTPUTS Goods, equipment or services that result in inputs that produce effects	OUTCOMES The project's effects on the direct beneficiaries resulting in the use of outputs	IMPACT Effects generated by the project at a general level
DAC concepts and definitions (OECD)	RESOURCES / ACTIVITIES Activities mobilising technical, financial and human resources to produce specific achievements	PRODUCTS Goods, equipment or services that result in activities that lead to direct effects	ACHIEVEMENTS / DIRECT EFFECTS What the action accomplishes in the short- and medium-term	IMPACT Positive and negative long-term effects produced by a development intervention, directly or indirectly, intended or unintended
Concepts used by SEO	RESOURCES / ACTIVITIES	IMMEDIATE RESULTS	SHORT/MEDIUM-TERM RESULTS	LONG-TERM RESULTS
Concepts and definitions of the intervention logic	INPUTS / ACTIVITIES Activities mobilising technical, financial and human resources to produce change	INTERMEDIATE RESULTS Products or services delivered	SPECIFIC OBJECTIVES Changes resulting from the use of outputs	OVERALL OBJECTIVE Contribution of the development action to changes at a more general level
BTC concepts		SPHERE OF CONTROL	SPHERE OF INFLUENCE	SPHERE OF INTEREST
		RESULTS products, achievements, impact of a development action		

The intervention logic approach was selected for the four impact evaluations. This means (i) that the definitions of outcomes and impact do not refer to the temporality of the effects; and (ii) that the projects' impact evaluation is structured around the logical framework by combining two methods: a quantitative method to assess and demonstrate the attribution of the effects on the direct beneficiaries³ (outcomes) where this is technically possible, and a qualitative method to assess the effect at a general level (impact) and assess the outcomes where the application of quantitative methods is problematic. This would be the same as saying that we are referring to theory based evaluation which uses both quantitative and qualitative methods.⁴

Figure 2 indicates that the term "results" means anything that is due to the inputs, i.e. outputs, outcomes and impact. This observation is particularly important when talking about results based management. Indeed, the M&E system should be able to assess a project's outputs, outcomes and impact.

2.2 One evaluation process, two methods

As mentioned above, **the methodology selected to evaluate the outcomes and impact is a combination of qualitative and quantitative methods structured around the intervention logic.** These methods are briefly presented below.

Qualitative methods give a reasoned judgement on the achievement of an intervention's impact, understood as a project's effect at a general level (ref. Figure 1 above). They analyse whether the conditions for achieving the desired impact are met at different stages of the intervention, in reference to the logical framework: design/formulation, input, implementation, completion and use of the outputs and achievement of outcomes and impact.

Qualitative analyses are based on information collected through documentary analysis, individual or group interviews and direct field observations. A good qualitative analysis provides an in-depth understanding of the context at the time of the intervention. In addition to situating the project in a historical context (previous and planned projects, etc.), this approach consists in understanding and identifying all socio-economic, political and cultural aspects, along with the power issues at different levels that may have influenced the project and/or the proper execution of the evaluation.

Quantitative methods measure and demonstrate the attribution of an intervention's outcomes, defined as the effects on the direct beneficiaries that can be attributed to the project. The objective is to establish a causal relationship by demonstrating empirically the extent to which a project - *and only this project* - has contributed to the effects observed on the beneficiaries (outcomes). This consists in using technical and econometric techniques on enough reliable data collected from the beneficiary group and a counterfactual, before and after the intervention (see below). This data can be collected in the field using household surveys and/or using secondary data, i.e. existing data from various sources (available through the projects' M&E system, national statistics, etc.).

It should be noted that the use of quantitative methods in project evaluation is generally part of a research perspective where the project's design is concurrent with that of the evaluation system, thus allowing the evaluation to be conducted under the best technical

³ A quantitative or qualitative analysis can also assess the effects of an intervention on indirect beneficiaries where relevant.

⁴ Theory based evaluation is often equated to qualitative methods because it is based on the analysis of the logical framework, which is little used by practitioners of quantitative methods (who often evaluate one aspect of a project from a research perspective and not the entire project). Theory based evaluation is not inconsistent with the use of quantitative methods, but it is not always systematically used with them.

conditions.⁵ Indeed, the feasibility of applying quantitative methods is highly dependent on the quantity and quality of the available data before and after the intervention and the possibility of identifying a credible counterfactual. There are many challenges and limits in applying these methods, particularly in an ex post evaluation context (see below).

Although qualitative analysis can support a reasoned judgement on the achievement of a project's outcomes, the application of quantitative methods will be favoured over qualitative analysis where this is technically possible. In fact, under ideal conditions, the strict application of quantitative methods can be used to quantify the outcomes and can scientifically demonstrate their causal relationship with the project (its attribution), with a certain degree of precision. However, if all the conditions needed for the proper application of the econometric tools are not met, the quantitative analysis cannot demonstrate attribution, but only a level of correlation between the observed effects and the project. The argument for the causal relationship will therefore have to be made using qualitative analysis. Furthermore, if the application of quantitative methods proves impossible due to the lack of data or even because of too great a bias, a rigorous qualitative analysis on the achievement of the outcomes is then required.

In terms of assessing an intervention's effect at a general level (impact), quantitative methods are ineffective, whereas qualitative methods are reliable. Conversely, the findings on the outcomes made using quantitative analysis can be used as strong arguments for assessing the impact and its scope. Sometimes these findings can be generalised or extrapolated if they are supported by a convincing explanation.

Although very different, these methods need each other to provide quality outcome and impact evaluations. Perfect understanding of the quantitative technique can never be a substitute for a rigorous analysis of an intervention's causal sequence, or a relevant argument made by experts with local knowledge (Deaton, 2012). Conversely, a convincing explanation for an intervention's effects cannot replace the measurements and scientific evidence of the attribution of a project's effects through the use of quantitative methods.

Finally, a quality outcome and impact analysis is not limited to identifying what has worked or not worked, but should also explain how and why these results were obtained or not obtained. A rigorous qualitative analysis of the context and the challenges at different levels, along with a good knowledge of the realities in the field are therefore necessary to explain the results obtained by applying statistical and/or econometric techniques.

In summary, the combination of qualitative and quantitative methods can provide a precise evaluation of the outcomes using a credible counterfactual as well as an impact evaluation through a rigorous factual analysis. In other words, a mixed methodology can be used to measure effects on the beneficiaries and demonstrate their attribution or correlation to the project (depending on the quantity of reliable data available and the possible use of a counterfactual for the proper application of quantitative methods). It can also provide a reasoned judgement on the achievement of a project's impact, based on certain scientifically proven observations through the quantitative analysis of the outcomes. Lastly, it is used to give an explanation of the process that has generated or inhibited a project's effects.

⁵ These methods were created and have been widely disseminated by J-Pal (Abdoul Latif Jamel Poverty Action lab). They have since been used by many university researchers to conduct impact evaluations on projects financed by various donors (World Bank, AFD, Dutch development cooperation, Norwegian development cooperation, etc.), but every time the projects to be evaluated have been formulated and implemented in line with an evaluation system, ensuring the best conditions for applying quantitative techniques. This exercise is different because these are ex post impact evaluations for which there is no impact evaluation system, therefore the application of quantitative methods is limited.

2.3 Quantitative methods: objectives and challenges

Quantitative methods are intended (1) to identify and measure the changes observed in the target population after the project's implementation; and (2) to establish, with a certain degree of precision, a causal relationship between the changes observed (expected or unexpected) and the project.

1) The quantitative methods' first objective is to measure a project's effects on the beneficiaries (outcomes).

Once the intervention's outcomes have been identified in the logical framework analysis, a series of questions must be asked:

- a. What are the measurable outcomes?
- b. How can they be measured?
- c. Is there data available to measure them?
- d. How can data be collected where there is none available?

Note that if there is no data and no way of collecting it or if an outcome cannot be measured, it is impossible to use quantitative methods. For example, in **Morocco**, measuring land productivity proved to be a relevant indicator, but its data collection was too complex with the available resources and time. This indicator was therefore abandoned. The outcome indicators for the three evaluations were created based on the field information collected, but not all of them could be subject to quantitative analyses, either because they proved to be unreliable or because there were insufficient observations. Data quantity and quality are two key elements for the application of quantitative methods (see below). Note that a qualitative variable can be processed using quantitative methods as long as it is systematically collected with a sufficiently large number of observations to be processed statistically or used in an econometric regression⁶.

2) The quantitative methods' second objective is to demonstrate the attribution of the effects to the intervention.

In order to conclude a causal relationship, the changes observed on the beneficiaries must be demonstrated to be as a result of the intervention.⁷ In other words, we must be able to assess what the situation would have been for the treated population in the absence of the project. It would therefore be ideal to compare the same individuals with or without the project's implementation, but by definition, this is impossible. Thus, one option is to compare the treated population with a population that has a significant number of similar characteristics prior to the intervention. This group is called the control group, comparison group or even the counterfactual.

Quantitative methods face several challenges in achieving these two objectives.

1. The choice of a good counterfactual

The treated group, also called the target group, is the beneficiary population selected to receive the development cooperation intervention. The counterfactual is a non-

⁶ Examples of qualitative questions that can be processed statistically: "Since the arrival of tap water in your village, do you think that your living conditions have improved?" YES/NO; "Since the arrival of tap water in your village, can you rate on a scale of 1 to 5, the welfare improvement for you and your family? (1: no improvement; 5: maximum improvement)" 1-2-3-4-5.

⁷ Sometimes, the econometric specifications possible with the data available present an endogeneity bias making the causal analysis complex or impossible. In these cases, the results obtained in the regressions are correlations between the analysed outcome and the project. There is no technical way to confirm that these effects are due to the intervention, they could be explained, at least in part, by unobserved, exogenous factors.

beneficiary population, not significantly different from the treated population before the intervention. In other words, the two groups are similar across a large number of characteristics, ideally, the only difference would be that the first are beneficiaries of the intervention and the second are not.

There are two types of evaluation design for determining the counterfactual: an experimental design and a quasi-experimental design.

- Considered to be the most rigorous of the evaluation methods, "**the experimental design**", also called "**the randomized control trial**" (RCT) is the first way to select a counterfactual. The basis of this method is the project's random allocation among a population of eligible beneficiaries to generate a treatment group and a statistically equivalent counterfactual before the intervention. This selection is often made at a certain level depending on the type of intervention, for example, at school or village level. When the selection is made at individual level, it raises certain ethical issues.⁸ With a large enough number of observations, the random allocation process can be used to create two groups whose average characteristics are statistically equivalent before the project begins. In theory, the use of random allocation therefore guarantees that these groups are equivalent. The collection of baseline data is used to verify this hypothesis empirically. If these two groups are indeed statistically equivalent before the project's implementation, then a single ex post difference between the average result recorded for the treated group and the average result observed for the control group will suffice.

The advantages of this design are: (i) the random nature of the choice of subjects which avoids the problem of selection bias (see below); (ii) the simplicity in interpreting the measurements of the effects attributed to the project (by the double difference method - see below for an explanation about the use of this method).

The main disadvantages are as follows: (i) it may be ethically or politically difficult to agree that one group and not another benefits from an intervention; (ii) a project's scope may not allow the creation of two groups (for example, a national project); (iii) the random nature may be difficult to respect or individuals from the two groups may move during the intervention (spillover effect, see below); and (iv) this is a costly method that requires a lot of time to collect and process a large amount of data.

- A **quasi-experimental** design means that the treatment group and control group are not determined randomly. It involves defining a counterfactual that presents statistically observable characteristics equivalent to the intervention's beneficiary group. This can be done before or after the project's implementation.

The advantages of such a design are that (i) it is often less costly and quicker than an experimental design, and (ii) it can be implemented for an ex post evaluation. The major disadvantage is that the statistical and econometric methods used to demonstrate attribution are often more complex, since it must be ensured that there has been no selection bias (see below). Thus the reliability of the results is sometimes called into question. Under- or over-estimates about the project's effects are possible.

There are two frequent problems that must be avoided in order to demonstrate that a good counterfactual has been found: the contamination problem or spill over effect, and the problem of selection bias of the beneficiaries.

⁸ See White (2013) "An introduction to the use of randomized control trials to evaluate development interventions", for more details on the application of RCT for project impact evaluations. See also Getler et al. (2011) "Impact Evaluation in Practice" for a simple explanation of the random allocation of groups.

- The two ways in which the spillover effect occurs are described below. It is possible to overcome this problem with an in-depth knowledge of the context in terms of both local dynamics and the presence of other donors in the field.
 - (a) An intervention has an indirect effect on the counterfactual. To ensure a certain similarity between both groups, there could be a tendency to take groups that are relatively close from a geographical perspective. However, there is a strong probability that a control group neighbouring the treated group could be indirectly affected by the project. Thus, although they are similar before the intervention, the post-intervention data collected from the control group is influenced by the project. This distorts the interpretation of the analyses. The spillover effect is an indirect effect of the project. Such an effect has been observed in the impact evaluation on access to tap water in Senegal where households in neighbouring non-beneficiary villages are now using the tap in the project's beneficiary villages.
 - (b) Effects of other projects on the counterfactual. It is possible for other donors to conduct the same type of intervention with the control group. It is also possible for different interventions to be conducted that affect the same outcomes. In both scenarios, a comparison in terms of outcomes between the target and control group will be biased.
- **Selection bias** appears when the reasons why a person participates in a project correspond to the results. This concept therefore refers to how individuals are selected to benefit from a project. If participation in the project is determined based on the beneficiaries' observable characteristics (gender, age, literacy, etc.), selection bias can be overcome via relatively simple econometric techniques. However, it is often argued that non-observable (or difficult to observe) characteristics influence beneficiaries' participation in a project (problem of self-selection). If these characteristics are correlated to the project's outcomes, then the analyses are biased (endogeneity bias). Indeed, the differences between the beneficiary and control group could be explained by the non-observable characteristics and not just by the fact of benefiting from the project.⁹

One way of overcoming this bias is the random allocation of the project within the population (randomized control trial, see above). Another way of avoiding this issue is to use the **pipeline** technique. This involves implementing a project sequentially and using the group of beneficiaries from an later execution phase as the counterfactual. However, the order in which the project is executed in the various intervention areas cannot be decided based on characteristics influenced by the project or influencing the project's outcomes, so it is preferable to choose a random order of execution where possible.

To demonstrate the quality of the chosen counterfactual, it is important to provide convincing justifications, limiting any suspicion of selection bias. The target group and counterfactual must be shown to be statistically similar before the intervention's implementation, hence the importance of baseline data for both groups. If this is technically not possible (lack of data), an argument based on good field knowledge that there is no selection bias, or at least the identification of the bias' direction (an over- or under-estimate of the effects) must nevertheless be made.

⁹ The following example illustrates this. In a microcredit project, the temptation is to compare individuals with and without microcredit in terms of generated income (outcome). However, it is highly probable that the most dynamic individuals would apply for access to microcredit. This characteristic, which is difficult to observe, determines access to the project and how the money grows. A comparison of individuals with and without microcredit would likely give an overestimate of the effect of access to microcredit.

The ex post evaluations of the projects in Morocco and Senegal are of quasi-experimental design insofar as, on the one hand no counterfactual was previously defined, and on the other hand, there is no baseline data available. The counterfactuals identified are not completely satisfactory, but the analysis recognises and takes this into account. In **Morocco**, the control villages have different pre-project characteristics to the target villages. These characteristics were probably observable (among other things, the type of agriculture) but were unfortunately not collected due to a lack of knowledge of the context. The interpretation of the results is therefore cautious and takes this bias into account as far as possible. Results that cannot be interpreted have been discarded. In **Senegal**, the treated and control villages also have pre-intervention differences. Due to the existence of participatory diagnostics that give a status of the situation in the villages before the intervention, some observable differences were able to be identified and incorporated into the analysis to identify the bias' direction (examples of observable characteristics at village level: population density, number of users per water point, isolation, etc.). There is also the issue of self-selection regarding private standpipe users. Access to a tap in the yard is an indirect effect of the project and the selection of beneficiaries with access to a private tap is still partly unknown due to a lack of observable variables collected, but also because of the potential influence of non-observable variables.

2. Use of enough quality data

The statistical and econometric methods used to assess an intervention's outcomes require enough reliable data. Having enough data means having a large enough number of observations to apply the statistical and econometric techniques (the more observations, the more accurate the tests), but also to be able to identify the effects, if any. Having quality data means that this data is relevant, that it has been clearly and precisely defined and that it has been carefully collected. The importance of defining an indicator and the many measurements and reporting biases that may exist when data is collected can never be over stressed. In the majority of evaluations, it is still a challenge to satisfy both conditions of quality and quantity, especially if quantitative methods are to be applied.

Data quantity/quality trade off. The higher the number of random observations is, the better the statistical analysis will be. In theory, there is a formula for determining the optimum sample size allowing confirmation with a certain degree of precision (power calculation) of whether or not there is an impact¹⁰. In practice¹¹, this formula remains in the background, since the sample size is calculated primarily according to the planned statistical-econometric method and time and budget constraints. However, there is no point having a large number of observations if the information collected is unreliable or useless. It is vital that data quality is not ignored.

There are two data types: secondary and primary data.

There are various sources of secondary data:

- project data (via the M&E system),
- data from national statistics,
- data collected within other projects,
- data available in certain public or private bodies (health records, school attendance registers, data from companies, etc.).

¹⁰ This formula depends on the sample size (the larger the sample, the greater the degree of precision), the intra cluster correlation, the minimum effect (the weaker the expected effect, the larger the sample needed), the targeted confidence interval (95% or 90% or 99%) and the survey costs. See White (2013) "An introduction to the use of randomized control trials to evaluate development interventions".

¹¹ I.e., when academic research perspective is not a priority.

In many countries it is a real challenge finding quality secondary data. In **Senegal**, the plan was to collect secondary data from schools and health centres (to assess the increase in school attendance and the reduction in diarrhoea), but the idea was quickly discarded, given the mixed quality of the existing data. Since health and school records are completed at the discretion of those in charge, they are not a reliable data source. In the **DRC**, the success rates for State examinations were collected through the school principals, along with the number of enrolments in the various subjects. In **Rwanda**, the case is somewhat unusual; secondary data of sufficient quality was available to study one outcome in particular (improved diagnosis of the thick smear test), but it was unable to be used due to a lack of authorisation from the Rwandan Ministry of Health.

When no secondary database can be used, or this is incomplete or unreliable, primary data must be collected.

Primary data is data that is collected directly by the evaluation team using detailed questionnaires sent to the appropriate respondents for the planned type of study. For example, households, one individual in the household, village chiefs, mayors, company bosses, hospital directors, school principals, etc. These surveys can be conducted at different levels for the same study. For example, in **Senegal**, a survey has been conducted at household level (217 households were questioned), as well as a survey at village level and among groups of borehole users. Although these last two types of survey were group interviews (focus group, see below), the data collected has been entered into the household database, in order to capture the differences between villages or borehole management groups. A household survey has also been conducted in **Morocco**. In the **DRC**, primary data has been collected at different levels through group interviews (focus groups) with principals, teachers, students and former graduate students. Given the small number of selected schools (two beneficiary schools and two control schools), no detailed statistical processing was conducted.

Primary data collection requires in-depth knowledge of the context and challenges in order to create the right collection tools. It is important to ask questions relevant to the themes studied in the local context, but these questions must also be well-formulated and use suitable vocabulary for the types of respondents. A primary data collection system requires significant mobilisation of field resources and involves the evaluators at every stage to ensure data reliability, from designing the questionnaires to supervising daily data collection in the field.

- **Focus group-type surveys** are structured/semi-structured interviews with larger or smaller groups of individuals from a single group (for example, village authorities, borehole managers, a group of teachers or even a group of students). The use of such a tool requires careful attention in order to identify inconsistencies in or the withholding of information by any of the participants. Although **this technique is a qualitative method**, it is mentioned here because information systematically collected in this way from different groups can be encoded in a database and then processed using statistical tools (as was the case for the project in **Senegal** where data collected through focus groups with the village authorities has been encoded).

It must be stressed that it is important for these groups to be selected at random for the collected data to have greater relevance. This also applies when certain individuals have to be selected in a group. In other words, in order for the evaluator to retain his independence, he must "go where he wants to go and not where it is suggested he goes". In addition, it must be ensured that no observer is able to inhibit any of the participants. For example in the **DRC**, there were no officials present during the interviews with the students or the teachers; in **Senegal**, there was no water distribution network manager present during the surveys with beneficiary or non-beneficiary households.

- Household surveys are interviews based on a detailed questionnaire administered by a researcher with one or several individuals in a household, a company, a school, etc. In the case of an evaluation, these individuals are selected at random

from a beneficiary and a non-beneficiary population (counterfactual). The benefit of this type of tool is that respondents feel more at ease than talking in a group interview. The data collected focuses on both quantitative (for example, how many people live in the household, what is the age of the head of the family) and qualitative information (for example, how satisfied are you with the use of the public taps, how useful is it to have a secondary diploma when looking for work in the DRC, etc.). These questionnaires also ask questions that must then be coded so that the responses can be processed statistically.

Implementing a household survey requires extensive logistics. A team of honest, hard-working enumerators must be created. The questionnaire must be tested to ensure the relevance of the questions and their understanding, for both respondents and enumerators. Then, during the survey period, enumerators need to be closely monitored to ensure the quality of the collected data and that it has been collected from the (randomly) pre-selected households. Entering this type of data also requires considerable supervision so as to ensure once again the reliability of the information encoded that will then be analysed.¹² Finally, the database must be "cleaned". This involves carrying out various tests to guarantee that the data has been encoded correctly and that answers are consistent. This means that there is no inconsistency between the answers to various questions in a questionnaire and all responses are consistent with field observations. For example, in the database for **Senegal**, an error in the computer checklist was not picked up by the survey supervisor which shifted all the answers in a section by one or two boxes. Everything had to be recoded. The consultants realised the mistake because some answers were inconsistent with their field observations, so the encoded data was checked with that on the original questionnaires to identify the source of the error. Without this check, the answers to one question could have been attributed to another question and so on.

This logistics involves a large amount of work. Working with a local partner who is an expert in quantitative evaluation techniques and has a network and experience in this type of undertaking facilitates the field work enormously.

The implementation of primary data collection generates a series of fixed costs (not dependent on the number of surveys or interviews to be conducted). For example, whatever the number of surveys to be conducted, a questionnaire must be drafted, enumerators and/or translators trained, field collection tools tested, corrections made, a computer input mask designed for encoding data, etc. Cutting corners during these crucial stages in the evaluation process inevitably leads to a loss of quality in the data to be analysed which may then challenge the relevance of the primary data collection (see above: quantity/quality trade-off).

Whatever type of data collection, those being observed may tend to change their behaviour because they know they are being observed. This is called "**the hawthorne effect**". This is a typical internal validity problem for impact evaluations. Respondents conforming to this "psychological" effect have a greater propensity to disguise reality (over- or under-estimate certain facts or data). It is therefore important to remain vigilant to the various challenges and strategies used by respondents during evaluations and try to pick up responses that could be under- or over-estimated. The quality of the questionnaire, the honesty of the enumerators and the time spent earning the trust of the local authorities and each respondent are just some of the elements that may greatly help to mitigate this effect. In **Morocco**, the local authorities probably influenced the respondents sometimes in one direction, sometimes in the other. Fortunately, due to a

¹² It should be noted that it is increasingly common to see surveys conducted using tablet computers where data entry is instant. However, the use of such a medium has some disadvantages such as the need for access to electricity, the need to find researchers familiar with computers, the difficulty of going back if an inconsistency is spotted during a survey, etc.

long-term field presence which helped to earn the trust of most of the respondents, this phenomenon was able to be identified, at least in part.

In order to demonstrate attribution, this **data must be available for several key periods in the intervention and collected from the beneficiary group and the counterfactual**. The ideal situation would be to have data for the pre-intervention period (**baseline**), during the intervention (**monitoring**) and the post-intervention period (**ex post**).

- **The absence of a baseline survey** may lead to a bias in the assessment of the outcomes: (i) the differences observed through an ex post survey between a target and control group may pre-date the project; and conversely, (ii) a lack of differences between these two groups may simply reflect a difference prior to the project, that has disappeared due to the intervention in the target group. An ex post comparison between participants and non-participants in a project also raises the issue of selection bias (see above: the reasons why someone participates in a programme correspond to the results).

Without a baseline, quantitative methods on ex post data can be used to compare the target and control group. The technique of propensity score matching is most commonly used in this case (see below for an explanation). However, without a baseline, the counterfactual's credibility must be proven by a rigorous qualitative argument and there is a greater risk of bias in the estimations (risk of comparing two incomparable groups, risk of not checking for non-observable variables influencing the result and the fact of belonging to a group).¹³ "Once the programme is implemented, if it is not possible to identify precisely how it has been attributed to the beneficiaries, and no baseline data has been collected, there is little, or even no possibility of a reliable quantitative evaluation" (Getler et al., 2011: p115).

When there is no baseline, where possible baseline data must be reconstructed using data from respondents' memories (recall data, see below) or reliable secondary data.

- **The absence of a counterfactual**. Without a comparison group, a project's effects are determined by examining changes in the results for the project participants over time. This involves making a before and after comparison of the beneficiaries under the hypothesis that without the project, the characteristics of the beneficiary group would not have changed. Unfortunately, in most cases, this hypothesis is invalid. Without a valid control group, it is therefore impossible to demonstrate empirically that the effects observed within a group of beneficiaries are the result of an intervention.
- Under **tight time and budget constraints**, a satisfactory option allowing quantitative methods to be applied is to collect baseline data from a sample of beneficiaries and reconstruct this data for the counterfactual through ex post recall data (see "Conducting impact evaluations under budget, time and data constraints", IEG, 2013).
- **Recall data** is data collected from individuals calling upon their memory to give the most accurate information possible about a situation prior to the survey period (for example, the pre-intervention period). Here, there has to be a time reference which is understood by everyone involved in the same way. However, the notion of time is very cultural so it is often useful to use historical references to take individuals back to the required situation (for example, during the election of such a president, at the time of the drought in such a year, before the construction of the public tap, etc.). Some information types lend themselves more readily to this technique than others. Indeed, exceptional data is more likely

¹³ See Getler et al. "Impact evaluation in Practice" p.114-115.

to be accurately remembered than information concerning daily facts. For example, in the case of **Morocco**, it is easier to remember the crops grown during the floods than to remember the amounts earned through farming. Indeed, flooding occurs on average every two years while farming takes place throughout the year, every year.

3. The rigorous application of statistical and econometric techniques.

Depending on the scenario, more or less sophisticated econometric models must be used to demonstrate empirically the causal relationship between observed effects and the development programme. The following elements influence the choice of technique: ex-ante or ex post design of the evaluation, random allocation of the intervention, the existence of baseline data, the existence or level of validity of a control group, data type, amount of data and its reliability. However, it is not enough to understand the theory. These models must be competently applied to actual databases and the results then interpreted not only in line with the hypotheses that underpin the econometric models but also in line with the realities in the field.

In the cases of **Morocco** and **Senegal**, there was no usable pre-intervention data. Even so, the team managed to apply quantitative methods with the aim of quantifying and demonstrating the attribution of certain outcomes. Two techniques were principally used on the data collected ex post from a group of beneficiaries and a counterfactual: **(i) the double difference method**, where the baseline data was reconstructed using recall data; and **(ii) the propensity score matching (PSM) technique** where matching is primarily done based on ex post data (which is not ideal, see above). The validity of the counterfactual is demonstrated empirically where possible, otherwise its validity is mainly argued based on observations.

Statistical analyses have also been performed: **difference of means or factor independence tests**. These different analysis techniques are briefly explained below (ref. "Handbook on impact evaluation, quantitative methods and practices", Khandker et al. (2010) and "Impact evaluation in Practice" Getler et al. (2011), for more details about these methods and practical applications).

❖ **Difference of means test**

This test is used to verify whether two groups are on average statistically different according to certain continuous characteristics (for example, the age of the head of the family, area of land owned, number of years of education, etc.).

The null hypothesis of the statistical test is that the two groups are on average statistically similar (H_0 : average of target group = average of control group). The "student" test is used to conclude whether the null hypothesis can be rejected (i.e. reject the fact that they are on average statistically similar) with a certain precision as to the chances of being wrong: 10%* chance of being wrong, 5%** or 1%*** chance of being wrong.¹ The more asterisks, the lower the chance of being wrong in rejecting the null hypothesis, i.e. the greater the confirmation that these groups are on average, different from each other.

Example:

Continuous characteristics	Target group	Control group	Significance
Average age of the head of the family	58.4 years	57.6 years	Not significant
Average number of years of education	10.8 years	9.5 years	**
Average surface area of land owned	9.13 ha	15.93 ha	***

Reading the results in this table:

- On average, the heads of the family of both groups are of similar ages.
- On average, the individuals of the target household have more than one additional year of education than individuals from the control group, this difference is significant at 5% (i.e. there is a 5% chance that this difference is not significant).

- On average, the target household has 5.8 ha less than the control household, this difference is significant at 1% (i.e. there is a 1% chance that this difference is not significant).

❖ **Factor independence test**

This chi-squared test tests the hypothesis that two discrete variables (binary or able to take several values) are independent of each other. One of the variables is dichotomous indicating whether the observations belong to the control group or a comparison group, the other represents a discrete characteristic, for example the fact of having a scooter, motorbike, cycle or car; the fact of belonging to a certain ethnic group (Wolof, Peulh, Serer).

The null hypothesis indicates that the variables are independent of each other. For example, the null hypothesis stipulates that the fact of belonging to a certain ethnic group is not related to the fact of being in the target group or control group. The number of asterisks indicates whether the fact that these variables are independent of each other can be rejected with a certain probability (see difference of means test, above). For example, the more asterisks, the more confidence that belonging to a certain ethnic group is related to being part of a (control or target) group.

Example:

Discrete variable: ethnic group	Target group	Control group	Significance
Wolof	60%	40%	***
Peulh	12%	29%	***
Serer	16%	22%	Not significant

Reading the results in this table:

- The proportion of Wolof households in the target group is significantly greater than in the control group. The two groups therefore differ significantly in this characteristic. Rejection of factor independence at 1% (i.e. there is a 1% chance of being wrong in confirming that there are significantly more Wolofs in the target group).
- The fact of belonging to the Serer ethnic group does not seem to be related to the fact of being in a certain group. There is thus no rejection of factor independence.

❖ **The double difference or difference in difference**

The effect of the treatment is determined by comparing the difference in terms of indicators of interest between the treated and control group before and after the project (the treatment) is implemented. A significant variation in the difference in terms of outcomes between the two groups over time is then attributed to the project.

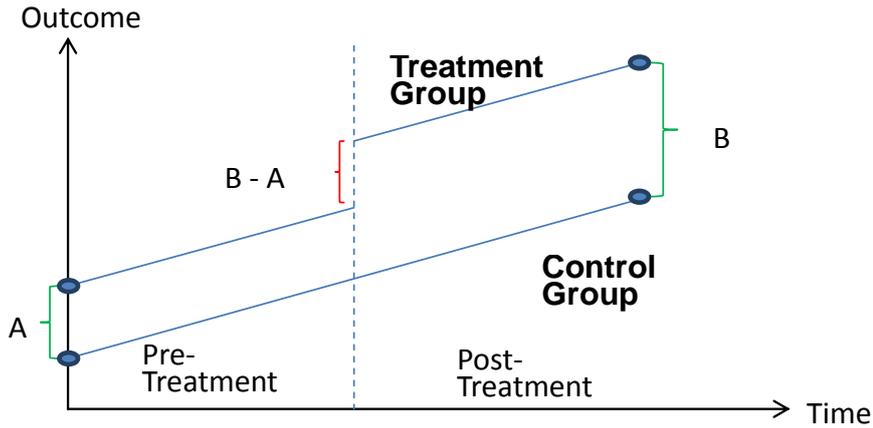
Step 1: Data is collected for a reference year before the project is implemented (or baseline data reconstructed using recall data with the appropriate caution regarding data reliability); and the difference between both groups (treated and control) before the programme is observed;

Step 2: Data is collected for a reference year after the project is implemented; and the difference between both groups (treated and control) after the project is observed;

Step 3: The variation in the difference between both groups (the method focuses on the difference between the differences) is observed. The difference in difference method is based on the following hypotheses:

- Non-observable characteristics do not vary over time: the selection bias, if there is one, is constant over time and non-observable characteristics that could influence the outcome do not change over time.
- Without treatment, both groups follow the same trend.

Figure 3: Diagram showing the double difference technique



Example:

Periods	Household type	Target households	Control households	Difference between household type in a period
Pre-project (2006)		132 minutes	150 minutes	18 minutes
Post-project (2012)		59 minutes	124 minutes	65 minutes ***
Pre/post project difference within the group		73 minutes ***	26 minutes	Difference in difference: 47 minutes ***

Reading the results in the table: Due to the treatment (the arrival of a public tap), target households experienced a time saving of 47 minutes a day.

❖ **Propensity score matching**

This method aims to compare a series of variables of interest (that measure the project's potential effect) between individuals who have the same probability of taking part in a project (this probability can be calculated in several ways).

Step 1: A valid counterfactual is created by matching the households based on observable characteristics (not influenced by the project) - ideally on baseline data-;

Step 2: Based on observable characteristics, the probability of participation in the treatment is estimated for all households;

Step 3: Each household in the treated group is compared to the households in the target group using a rule to be defined. For example comparison to the 5 closest "neighbours" in terms of the probability of taking part (calculation method used in the Morocco evaluation);

Step 4: The benefit due to the treatment is calculated through the difference between the outcome for the treated household and the average outcome for its 5 closest "neighbours" (or according to the rule chosen at step 3);

Step 5: The result obtained by this method is the total average benefit of the treatment. It is given by the average of the individual benefits calculated at step 4.

The PSM method is based on the **hypothesis** that there are no non-observable characteristics likely to explain both the studied effect of the project and the fact of being in the treated group (no selection bias). The major problem with this method is that it is impossible to prove empirically that there are no non-observable characteristics likely to influence participation in the project and the result. In addition, this technique requires the collection of large databases on significant samples. Furthermore, there is always a risk of a lack of mutual support between the two groups (target and control). It should also be noted that

matching on ex post data is not without risk. A match can unwittingly be made based on characteristics that have been affected by the aid programme.

Example:

Variables	Coefficient	Significance	No. of obs.
Average share of farming income in total income	14	*** (1%)	218

Reading the results in this table:

The PSM method concludes that the treatment has had a significant effect (at 1%) on the proportion of total income from farming. This effect is positive. The proportion of farming income is almost 40% of total income for the treated group compared with 26% for the control group (a difference of 14 percentage points is the coefficient given by the estimate).

4. **Correct and consistent interpretation of the results**

Although it is easy to read the results once the quantitative techniques have been mastered, interpretation gives rise to some difficult questions. On the one hand, the application of two different models cannot give the same results. Therefore, either the analyses must continue in order to find a robust result¹⁴, or we must remain cautious regarding the existence of this result because it is not robust. On the other hand, the results obtained may contradict field observations. So, we must attempt to understand this result by ensuring that the methods have been properly applied, that the data is correct and that the field observations are convincing. If there is a discrepancy between a quantitative result and a qualitative analysis, this should be flagged and this result interpreted with caution.

The conclusion on whether or not an outcome is related to a development intervention must also be treated with caution. Indeed, several elements may cause bias in the results obtained. Firstly, a small sample size may reduce the precision of the estimates making it more difficult to identify any real effect. Secondly, the proper use of econometric methods must be ensured. To do so, estimates must be made that make sense both technically and contextually and which take into account any potential bias (selection bias, contagion bias). Finally, data reliability is essential; as much as possible must be done to protect against measurement errors and strategies by respondents to manipulate the answers. Furthermore, readers must always be critical as to the potential manipulation of figures by practitioners ("How to lie with statistics", Darrell, 1993). Reproduction of field results is often a good way to avoid this kind of manipulation and thus any misinterpretation of the empirical results.

Interpretation of the results therefore requires good technical knowledge and intellectual honesty but also a good understanding of the project's theme and the local context. Consequently, it is easier for an evaluator to interpret the results of a study when he has managed the data collection process from start to finish and been responsible for the analysis. This was the case for the two ex post impact evaluations conducted in **Senegal** and **Morocco**: the CRED researcher was present throughout the evaluation process.

In the case of **Morocco**, the only study where both techniques have been tested, the team was confronted with the fact that these two techniques did not always produce the same result. Obviously, this means that the results are not robust (reliable), but it also illustrates the inherent difficulty in interpreting the results, the consideration of potential bias and the honesty and caution required in interpreting the results.

¹⁴ A robust result means a result that is maintained across different models and under different specifications. In practical terms, this means that a result is significant, of relatively the same size and same direction (positive or negative) when it is applied for example to a difference in difference model and a PSM model and when for the same model, some explanatory variables are changed.

2.4 The evaluation's practical organisation

Published impact evaluations that use mixed methods have often been performed by teams specialised in one of these methods. They have been conducted:

- Either by practitioners of qualitative methods who use quantitative methods at certain times. However, in view of the complexity of implementing quantitative methods, these evaluations are often lacking in rigour on several levels.
- Or by practitioners of quantitative methods (primarily individuals from the world of academic research) who use qualitative methods to justify certain methodological choices and interpret the results. However, these individuals often view qualitative methods as second-best.

The specific feature of the impact evaluations conducted here lies in the fact that each evaluation was carried out by two teams working in parallel, each specialised in one type of method: ADE for the application of qualitative methods and the CRED for the application of quantitative methods.

The case of the **DRC** is slightly unusual because the evaluation process used only qualitative methods but was based on a quantitative evaluation process. Indeed, this was a case study based on individual and group interviews. However, the choice of the province and the respondents was made as randomly as possible and special attention was given to questioning direct project beneficiaries. In addition, an approach with a counterfactual was selected and the time spent in the field was longer than for a standard qualitative study.

Although the evaluation processes were implemented relatively independently by both teams, they produced a joint evaluation report for each evaluation. The idea was to enhance both sets of findings on outcomes and impact without *a priori* defining them in the evaluation process to be followed. There were a wealth of lessons in pooling the results to produce a single quality report per project and holding discussions with the steering committee.

Figure 4 below shows the **key steps followed by each of the teams, then those for the joint work**.

Figure 4: Key steps in the evaluation process by mixed methods

	ADE « qualitative methods »	CRED « quantitative methods »
Preparatory phase	Reconstruction of the project's intervention logic	Identification of the realistic and measurable outcomes Identification of the methodological strategy Exploratory mission
Field phase	Comparison of the causal relationship with field realities	Primary data collection Data entry
Analysis phase	Assessment of the impact based on reasoned judgement	Data cleaning and processing Outcomes measures and demonstration of their attribution

Exchange between the teams	
Step 1	Comparison of the reconstructed intervention logics
Step 2	Comparison of the outcomes and impact assessments; Explanation of the underlying mechanisms
Step 3	Writing a joint evaluation report

For the **ADE team**, specialising in the application of qualitative methods, the sequence of steps was as follows:

- **Preparatory phase: Reconstruction of the project's intervention logic** based on a documentary analysis and a field mission. This work is used to highlight the anticipated sequence of results which, through causal relationships, should produce the expected impact.
- **Field phase: Comparison of the causal relationships with realities in the field**, using observations of the achievements and their use, expert opinions, semi-structured interviews with various stakeholders both in Belgium and the intervention countries (including some beneficiaries). The experts also relied on an analysis with a counterfactual although this was not statistically representative (this is the case in the **DRC**).
- **Analysis phase: Assessment of the impact at a general level** based on a reasoned judgement of the intervention's contribution to the impact (on a general level, see definition above). Qualitative analysis methods are generally recognised as ineffective for precisely quantifying the impact and its attribution to the project. However, their strength lies in their ability to clarify the project's potential contribution to the impact in complex contexts.

The evaluation process for the **CRED team**, specialising in the application of quantitative methods, is briefly described below:

- **Preparatory phase:**
 - **Identification of the intervention's realistic and measurable outcomes** based on reading the available project documents but also existing academic literature on the themes addressed in the project and other impact evaluations conducted in the same areas/sectors.
 - **Identification of the methodological strategy**. This consists in identifying the appropriate quantitative methods to quantify and attribute the effects (outcomes) to the project and identifying a data collection strategy after drawing up a list of existing data that could be used for a quantitative analysis. This step involves defining the sample, the counterfactual, the data collection tools to be used, etc.
 - **Design of primary data collection tools** (household questionnaire, interview guide for the focus groups) through a good knowledge of the context as well as the project and the outcomes to be measured.
 - **Exploratory mission**. This mission has several objectives.
 - (i) It aims to provide a better understanding of the context, challenges and realities in the field by meeting with various key individuals from the project and the intervention sector and visiting the intervention area.
 - (ii) It is used to verify a certain number of points that are crucial to the analysis. This includes verifying the relevance of the choice of outcomes to be measured. Consequently, this must start by observing the achievement of the outputs. It is also essential to check the reliability of the available secondary data and the feasibility of the planned primary data collection strategy.
 - (iii) This is the opportunity to meet the local partner and organise the logistics for the field phase (creation of a team of enumerators, financial arrangements, field movements, etc.) and refine the methodology based on the partner's field knowledge.
 - (iv) Finally, it is also the time to test the survey tools and make any corrections.

- **Field phase:**
 - **Primary data collection.** This step begins by training the enumerators and testing the questionnaire with a sample of beneficiaries and non-beneficiaries. Next the field surveys take place closely supervised by the evaluator to guarantee the reliability of the data collected.
 - **Data entry** into computer format (for example using the CSPro programme), supervised by the local partner (data encoding is a slow and therefore costly process in our country, so it is preferable to perform this on site).
- **Analysis phase:**
 - **Data cleaning** to verify the quality of the encoding (data entry consistent with data on the paper questionnaires).
 - **Data processing by applying quantitative methods** (for example using the software package STATA). The hypotheses underpinning the project's effects are tested using statistical and econometric techniques.
 - Based on the analysis results, **the team makes a decision on the achievement of the outcomes, quantifies them and demonstrates their attribution**, where appropriate.

At the end of the process, results are being exchanged between the two teams.

- **Step 1: Comparison of the intervention logic** explicitly reconstructed by ADE with the intervention logic implicitly reconstructed by the CRED during the identification of the outcomes to be assessed.¹⁵
- **Step 2: Comparison of the outcome and impact evaluations** for each project. At this stage, several elements discovered using one or the other method are used as additional arguments to the findings made by each team. Some findings are also challenged. This is where the mixed method is of critical importance.
- **Step 3: Writing a joint evaluation report.**

¹⁵ The CRED researchers, like many academics do not use the intervention logic. However, in order to identify the measurable outcomes, they have unknowingly reconstructed one. This is why the word "implicitly" is used. In the future, this step should be completed jointly, in order to save time.

3. Main conclusions on the achievement of the outcomes and impact of the 3 projects

Firstly, this section presents a summary of conclusions about the achievement of the outcomes and impact of the three projects. Secondly, a summary of the three projects along with the methodology used and the results obtained are presented in table 1 below. For more information about the projects and their evaluation, the reader is invited to read the project reports published in the appendices.

The Rwanda project is not included because the CRED was unable, legally, to collect all the secondary data necessary for the quantitative analysis, despite complying with the process required by the Ministry of Health. Quantitative analysis through a scientific research protocol has been presented with a view to obtaining the approval of the National Health Research Committee.

3.1 Evaluation of a Project supporting Technical and Vocational Education in the DRC (AETP2)

The methodology selected for this evaluation was **a case study using a counterfactual in the city of Kisangani**. This study was therefore based on a rigorous qualitative analysis of field observations and information collected through individual and group interviews in the beneficiary and control group schools along with key individuals involved in this sector and/or working on the project in Belgium and the DRC. In-depth interviews were conducted with school principals, teachers, students and graduates. Given the lack of usable quantitative data for this project and the fact that implementing a primary data collection system proved irrelevant following the findings of the exploratory mission, the application of quantitative methods had to be abandoned.

The exercise demonstrates that the **AETP2 project did not have any tangible effect on the project beneficiaries in the city of Kisangani**. The quality of the education has not improved; the number of staff has not increased; the schools have no capacity for self-financing; and the young graduates are no more employable. This failure can in part be explained by the very partial use of outputs, but particularly by the extremely unfavourable environment in this country's technical and vocational education sector. On the one hand, for many decades, the country has been confronted with failing infrastructure, aggravated by violent conflicts. On the other hand, a lack of appropriate incentive systems and even the existence of adverse incentives are harmful to the effectiveness of any action aimed at improving the quality of education. In this regard, the report is a wealth of extremely precise and detailed information.

Furthermore, there are elements that lead us to believe that these findings are not specific to the city of Kisangani, because the institutional context into which the project fits applies across the entire country. In addition, the intervention logic and the M&E method of the BTC are identical in all provinces. This means that even if the achievement of the outputs is better in some schools, the impact of the AETP2 project is certainly no more significant elsewhere (except perhaps for a few exceptions related primarily to the personalities of those involved).

An extrapolation of the findings of the case study is proposed, to support a reasoned judgement on the low achievement of the outcomes across all beneficiary schools and the lack of a general impact for this project. It should be noted that even if the outcomes had been achieved, the impact on a national or even provincial scale would not have been significant, because the AETP 1&2 projects only concerned 5% of the country's technical and vocational schools.

Unfortunately, the principal lesson to be learned is that under the conditions highlighted, the programme to support technical and vocational education should never have taken place or should have been designed in a different way. It is indeed impossible to support effective education when the State itself has not established an adequate policy supported by appropriate budgetary resources. **In the DRC, far from being supportive the education policy is discriminating.** Firstly, the State only pays for part of the teaching staff (teachers paid by the government). Secondly, the salaries paid are clearly insufficient to ensure a decent standard of living, which forces students' parents to top up teachers' salaries. So for households the cost of school is very high, because these "motivation" contributions must be added to a series of other costs which include money not only for books and uniforms but also for the consumables required to operate the schools' equipment. Finally, in addition to setting aside derisory amounts to fund education, the Congolese State imposes a tax on parental contributions to teachers' salaries. These difficulties come on top of all the problems caused by the country's failing infrastructure, particularly the frequent lack of electricity that compromises the operation of technical equipment.

These observations reflect a superficial knowledge of the sector and **management that is more focused on achieving outputs than the desired outcomes and impact.**

Although the report's assessment is based on a simple type of methodology (a case study with counterfactual in the city of Kisangani), **it is able to highlight the key aspects of the failure.** There is enough support for these factors inhibiting any effect of such an intervention, to convince that the conclusion would undoubtedly still be valid if a more sophisticated methodology had been conducted under satisfactory conditions. The use of a counterfactual does not add a great deal to the strength of the diagnosis, but this could not have been discovered prior to the field work. The counterfactual was used to verify the selection criteria for the schools established by the project and to triangulate certain information as well as providing a better understanding of the context and challenges at different levels.

3.2 Evaluation of a Project for the Improvement and Reinforcement of Water Points in Senegal (PARPEBA)

Due to the lack of usable baseline data, a quasi-experimental design was implemented to collect **primary data from a target group** (beneficiaries with access to tap water) **and a control group** (direct non-beneficiaries because they do not have access to tap water in their village). The analyses conducted, sometimes through the application of quantitative methods (double difference), sometimes using a qualitative argument, go into many details and contain a wealth of lessons for the donor and local managers. This abundance of information owes much to the fact that the analysis was based on household surveys and the relatively long-term field presence of the evaluators. This made it possible to examine many aspects of the behaviours and attitudes of rural water users in greater depth.

Although the positive effects of access to tap water were demonstrated, **water quality in terms of both bacterial and chemical contamination continues to be a real problem.** Indeed, over two thirds of the sites that were constructed or rehabilitated through PARPEBA have harmful fluorine and salt levels. In addition, there are many water contamination sources between the point of supply and point of use. The population is generally unaware of these water quality problems. Such an observation

calls for urgent corrective measures to prevent the risk of serious health problems among the rural population in the near future. Apart from the enormous problem of water quality, overall the project had **a positive effect on the beneficiaries and to a lesser extent the non-beneficiaries**. Indeed, the non-beneficiaries also use tap water, although they do not have direct access to it in their village.

The study shows that the use of tap water has resulted in **significant time and energy savings, especially for those beneficiaries with access to a private tap** (self-financed). The time and energy saved did not seem to have an effect on household water consumption. However, these time and energy savings seem to have encouraged people not only to start up small businesses, especially during the dry season, but also to participate in local groups and associations. These last two effects may obviously be related. A more detailed survey would be necessary to clarify this issue insofar as the survey did not detect any increased income among beneficiary households.

Another effect of the new boreholes highlighted by the microeconomic survey is **an increase in well-being and improved social cohesion** (a reduction in conflicts between women, who are primarily responsible for collecting water). It should be noted that this result is measured based on opinions given and not on factual observations, and must therefore be viewed with caution since there could be a response bias. In fact, the possibility of users wanting to give a positive image of the project's effects in a survey manifestly designed to evaluate its impact cannot be ruled out. It would be worth studying the effect on education in greater detail, particularly a reduction in the gender gap in school attendance, because there are many exogenous factors that could explain the observed trend. One of the great merits of the household surveys conducted in Senegal is that they highlight a certain number of unexpected indirect effects of enterprise cooperation.

It is impossible to quantify the contribution of the PARPEBA project to increased access to water in the Arachidier basin region because the baseline is confused and the calculation method to be performed is unclear in the project and sector literature. However, a very rough estimate showed that **PARPEBA helped to offer improved water access to at least 15.5% of the population in the affected region**. Unfortunately, this percentage is divided by 3, if only boreholes offering water with a good chemical composition are taken into consideration.

3.3 Impact evaluation of a Project for Small and Medium Hydraulics in Morocco (PMH)

Due to the lack of usable baseline data, a **quasi-experimental design was also used to compare a target group** (farmers benefiting from the rehabilitation of the traditional flood irrigation system) **with a control group** (farmers who did not benefit from this type of rehabilitation). The data collected through household surveys was subject to two statistical methods: double difference tests through the reconstruction of some baseline data using respondents' memories; and the use of propensity score matching mostly performed on post-intervention data.

The analysis shows that as a result of the operations to rehabilitate the flood irrigation system, **the project had a positive impact on the surface area of cultivated land during periods of water stress**, although no improvement was observed during flood periods. Furthermore, the project does not seem to have encouraged either crop diversification or changes in farming practices. In other words, it would seem that while the project did not result in a reduction in the interannual variation in agricultural yields (hence the persistence of diversification strategies based on migration), it nevertheless succeeded in **increasing the average agricultural income of farmers**. Compared with those farmers who did not benefit from the project, the beneficiaries derive most of their income from farming activities during both dry and flood periods. This last conclusion should be treated with caution as it is based on an income assessment by the respondents themselves. Observations on yields (not available in the collected data)

3. Main conclusions on the achievement of the outcomes and impact of the 3 projects

must be spread over several years in order to be able to conclude with any certainty that the project has helped raise the expected income of the farmers concerned.

The project also encouraged the spontaneous extension of the irrigation networks, especially upstream. Furthermore, it has potentially allowed the groundwater system to be recharged.

It would clearly be worth monitoring and evaluating this project over several successive years since despite its weaknesses, it is of great value. Indeed, although we need to remain cautious in the interpretation of the results, it would seem that the farmers dependent on this type of irrigation have seen their standard of living improve noticeably due to the project. Given that the PMH project affects no less than 70% of the land with flood irrigation in the province of Tiznit, we note **a generally positive impact**.

Table 1: Summary presentation of the 3 evaluation reports.

	Project supporting technical and vocational education in the DRC (AETP2)	Project for the improvement and reinforcement of water points in the Arachidier Basin in Senegal (PARPEBA).	Project for small and medium hydraulics in the Tiznit region in Morocco (PMH)
PROJECT DATA			
Years and budget	2005-2008 / 4,100,000 Euros	2003-2008 / 14,976,444 Euros	2004-2007 / 960,000 Euros
Geographic area	8 provinces	The Arachidier Basin (3 regions)	The Tiznit Plain (3 municipalities)
Sectoral policy	Non-existent.	Existing. Access to water is more important than water quality.	Existing. Beneficiary participation is denied.
Main activities	<ul style="list-style-type: none"> - Rehabilitation of 27 workshops for 6 promising sectors in 21 schools - Delivery of technical and teaching equipment - Teacher training - Raise awareness of TVE - Encourage IGI within schools - Creation of training/job units within schools 	<ul style="list-style-type: none"> - Creation of 32 multi-village boreholes - Rehabilitation of 14 boreholes - Creation and strengthening of the abilities of ASUFOR - Water purification techniques (1 inverse osmosis station + 1 test village with calcined bone) - Sanitation component (14 public toilets + hygiene training for 2 women per ASUFOR) 	<ul style="list-style-type: none"> - Rehabilitation of 13 flood barriers and 27km of channels - Protection of 22 km of channels (planting on slopes, mechanical protection of structures, etc.) - Distribution and planting of 52,000 fruit trees - Training in fruit production and grafting carobs - Training in administrative and financial management for AUEA
Beneficiaries	Schools, teachers, students	Rural population	Farmers
Expected outcomes	<ul style="list-style-type: none"> (1) Attractiveness of TVE in the supported sectors for the assisted schools; (2) Improvement in the quality of TVE in the supported schools; (3) Income creation and self-financing capacity for AETP2 schools; (4) Employability of young graduates from the supported sectors/schools. 	<ul style="list-style-type: none"> (1) Access to an improved water source; (2) Sustainable access to water; (3) Access to enough water at an affordable price; (4) Access to water of a satisfactory quality. <p>→ Save time and energy, improve the socio-economic and health situation</p>	<ul style="list-style-type: none"> (1) Increase irrigable areas (2) Increase cultivated areas (3) Crop diversification (4) Intensify agricultural activity: more crops per growing season (5) Better management of water resources and the irrigation system <p>→ Save time and develop new activities</p>
Expected impacts	<ul style="list-style-type: none"> - Improvement of TVE in the DRC (Involved in the achievement of MDG 3 - education for all) 	<ul style="list-style-type: none"> - Sustainability of water supply systems - Involved in MDG 7 (coverage rates for access to drinking water) - Improvement of living conditions for rural populations in the area 	<ul style="list-style-type: none"> - Guaranteed and stable income for farmers - Reduction in the rural exodus - Improvement of living and working conditions for rural populations

3. Main conclusions on the achievement of the outcomes and impact of the 3 projects

	Project supporting technical and vocational education in the DRC (AETP2)	Project for the improvement and reinforcement of water points in the Arachidier Basin in Senegal (PARPEBA).	Project for small and medium hydraulics in the Tiznit region in Morocco (PMH)
METHODOLOGICAL APPROACH			
Usable pre-project data? (baseline)	NO	In theory, yes, but not accessible in practice.	NO
Data collected by the ME system is usable?	NO, ME system focused on the activities implemented (outputs).	NO, ME system focused on operational monitoring (outputs).	NO, ME system focused on operational monitoring (outputs).
Exploratory mission	YES in Kinshasa: test of survey tools in Kinshasa (AETP1 target and control schools), logistics organisation, interviews with key people DGD/BTC/ministers	YES in the Arachidier Basin: focus group with ASUFOR, logistics organisation for the household survey, identification of the counterfactual, test of village and household questionnaire	YES in the Tiznit Plain: identification of a counterfactual, creation of survey frames and samples, logistics organisation for the household survey, test of household questionnaire
Data collection tools	<p>Interviews with key people in Brussels and Kinshasa</p> <p>Case study with counterfactual in the city of Kisangani: 2 target schools / 2 control schools; 4 classes (2 per school).</p> <ul style="list-style-type: none"> - Interview with 4 school principals - 2 focus groups with teachers - 3 focus groups with year 7 pupils - 1 focus group with graduates - Collection of secondary statistical data 	<p>Interviews with key people in Brussels, Dakar and the three regions affected by the project</p> <p>Household survey + focus groups with counterfactual in the region of Kaolack, new boreholes: 2 boreholes offering good quality water, 2 others with high fluorine and salt levels;</p> <ul style="list-style-type: none"> - 4 focus groups with ASUFOR - 30 focus group with village authorities - 217 household surveys 	<p>Interviews with key people in Brussels, Rabah and Tiznit</p> <p>Household survey with counterfactual in the Tiznit Plain (6 municipalities affected: 3 target group and 3 counterfactual): 9 areas having benefited from the project and 5 villages where the traditional irrigation system remains</p> <ul style="list-style-type: none"> - 231 surveys: 129 in the target group and 102 in the counterfactual
Analysis tools	<ul style="list-style-type: none"> - Reconstruction of the intervention logic - Secondary data statistical analysis - Qualitative comparisons of the different target and control groups 	<ul style="list-style-type: none"> - Reconstruction of the intervention logic - Difference of means or factor independence tests on primary data (household and individual level) - Double difference tests 	<ul style="list-style-type: none"> - Reconstruction of the intervention logic - Difference of means or factor independence tests on primary data - Econometric estimates: difference in difference model and propensity score matching technique

3. Main conclusions on the achievement of the outcomes and impact of the 3 projects

	Project supporting technical and vocational education in the DRC (AETP2)	Project for the improvement and reinforcement of water points in the Arachidier Basin in Senegal (PARPEBA).	Project for small and medium hydraulics in the Tiznit region in Morocco (PMH)
PRINCIPAL FINDINGS ON THE ACHIEVEMENT OF OUTCOMES AND IMPACT			
Outcomes	<ul style="list-style-type: none"> - Few significant effects on the beneficiaries since little or no use of the outputs - Extension of the ability to teach all practical lessons within the supported schools due to the renewal of old equipment and the delivery of small tools - Temporary improvement in the quality of the education due to the use of consumables delivered to the supported schools (now non-existent) - Possible extrapolation of the conclusions from the case study to all affected areas 	<ul style="list-style-type: none"> - PARPEBA helped to provide <u>improved and sustainable access to water</u> for the population of the connected villages but also, to a lesser extent, to the population of neighbouring unconnected villages - Field observations seem to indicate that PARPEBA did not result in an increase in water quantities consumed per day and per person - Significant time and energy savings (especially if access to a private tap) - Highly likely reallocation of this time and energy to IGI, well-being and social cohesion (perhaps also to the education of girls) - Strong inclination to pay for tap water - 2/3 of boreholes offer poor quality water, long-term dangers for health - Unconvincing sanitation and water purification components - Other unexpected positive and negative effects (see report) 	<ul style="list-style-type: none"> - The PMH project has not had an effect on the ability to irrigate during flood periods However, the project had a <u>positive effect in periods of low flooding</u> (larger surface areas irrigated among the beneficiaries) - Field observations show that the PMH project does not seem to have had an effect on the ability of beneficiary households to cultivate a larger proportion of their irrigated land - The project does not seem to have had a positive effect on maintaining or abandoning farming activities - The analyses do not provide a conclusion on whether the project had the effect of increasing the number of crops per growing season for beneficiaries. -The beneficiaries of the PMH project derive most of their income from farming activities during both dry and flood periods - No evidence that the project contributed to crop diversification
Impact	<ul style="list-style-type: none"> - No impact on improving the quality of TVE in the DRC - Nationally, the AETP 1&2 projects affect 5% of schools, thus there could not be a significant national impact, even if there was an impact at the level of the supported schools - All Belgian interventions in the sector contributed, at least indirectly to the signs of improvement in the TVE (different reforms are underway) 	<ul style="list-style-type: none"> - PARPEBA made a significant and sustainable contribution to improving the living conditions of an appreciable part of the population in the three regions concerned - But there is a great risk to the health of the population because 2/3 of the boreholes provide unsafe drinking water - Difficulties in assessing its contribution to MDG7 due to a confused baseline as well as the measurement of the indicators to be considered 	<ul style="list-style-type: none"> - The PMH project made a significant contribution insofar as the project affected almost 70% of land with flood irrigation in the Province of Tiznit - No evidence as to the reduction in the rural exodus - No evidence in terms of the notable improvement in beneficiaries' living conditions and at a more general level, the region's social conditions - Two unexpected impacts: a continuation in the extension of the networks (mainly upstream); and the project may have a positive effect on recharging the groundwater system.

4. Lessons from the use of mixed methods for evaluating outcomes and impact

This section firstly presents the methodology-type lessons on the evaluation process as a whole. It then examines the lessons on the design, implementation and monitoring-evaluation of interventions. Indeed, impact evaluation raises questions about how the project cycle works.

4.1 Lessons on the evaluation process

The application of mixed methods based on the intervention logic is a suitable solution for assessing and explaining a project's outcomes and impact.

Theory based evaluation is a comprehensive approach that can use both quantitative and qualitative methods to assess a project's outcomes and impact.

The **evaluative process combining qualitative and quantitative methods in the logical framework analysis** was used to:

- **Measure and demonstrate the attribution of certain outcomes of the evaluated projects.** Indeed, due to a lack of baseline data and the difficulty in using reliable recall data to reconstruct the project's previous indicators, only a minority of outcomes attributable to the projects in Senegal and Morocco¹⁶ could be analysed by the application of quantitative methods.
- **Assess and present the arguments on the attribution of all other outcomes that cannot be processed quantitatively through a rigorous qualitative analysis.**
- **Support a reasoned and rigorous judgement on the achievement of the impact** of the evaluated projects;
- Demonstrate the mechanisms and context elements explaining the observed effects or non-effects. The evaluations cast new light on the factors inhibiting or promoting the achievement of the outcomes and impact. **Understanding the "why" and "how" of a project's effects or lack of effects** is a key aspect of an impact evaluation.

None of the ex post evaluations conducted represented an ideal case for the application of quantitative methods: no baseline available, imperfect counterfactuals, relatively small samples, tight time constraints.¹⁷

¹⁶ Household surveys were used to evaluate the projects in Senegal and Morocco, which allowed quantitative techniques to be used.

¹⁷ See "Conducting impact evaluation under budget, time and data constraints" (IEG, 2013), for explanations about the methodological decisions to be made in such situations.

However, this does not mean that a quantitative evaluation was abandoned. Ultimately, it provides a different perspective on the achievement of the outcomes, even if for just a minority, and helps increase the rigour of the qualitative analyses. In addition, it highlights that only a quantitative analysis can demonstrate the attribution of the project's effects. This issue of attribution is central to impact evaluation.

The exercise conducted shows that when assessing a project's outcomes and impact, the issue is not the choice of one method or another but the **right combination of qualitative and quantitative methods**. Indeed, when they can be applied, quantitative methods can objectify the measurement and attribution of an intervention's effects on the beneficiaries. In other words, they are used to demonstrate the achievement of the outcomes empirically. However, they do not issue conclusions on the achievement of the impact (effect of the project on a general level, see definition above). At most they are used as an argument to support the qualitative impact analysis through the extrapolation or generalisation of certain empirically demonstrated findings. As for the qualitative methods, these set out a reasoned judgement on the achievement of the outcomes that cannot be processed quantitatively and on the achievement of the impact (effects of the projects at a general level). Finally, an explanation of the achievement or failure of the desired results has been proposed due to a thorough knowledge of the context and project gained by implementing the evaluative process combining both methods (see below).

Although the literature and these experiences show the relevance of combining quantitative methods with qualitative methods to assess a project's effects, it does not mean that it is easy and appropriate for any type of intervention.

While qualitative methods are almost always applicable, applying quantitative methods requires additional conditions: enough reliable data for the relevant indicators measured before and after the project with a beneficiary group and a credible counterfactual. **In addition to enough reliable and relevant data, the application of mixed methods requires time.** These two conditions are far from the realities of evaluations, where "*there is a need for speed*", and far from the realities of Belgian bilateral cooperation projects since "*very little usable data is available*".

However, not making any use of quantitative methods would mean that we would not have any objective measurement of the effects on the beneficiaries and no scientific evidence of their attribution to the project. It would also mean that we would simply ignore a large number of usable qualitative observations that can build a rigorous argument on the achievement of the non-quantifiable outcomes and impact. Thus, the feasibility of applying both methodologies must be decided in advance.

The exercise carried out did not try to compare the methods of analysis, as is often the case in the literature. Indeed, on the one hand, quantitative methods are not feasible without a qualitative dimension. On the other hand, qualitative methods cannot be a substitute for quantitative methods insofar as they do not answer the same questions: they are unable to measure and demonstrate outcomes empirically. However, it is essential to understand their specific features and synergies in order to make best use of them in an impact evaluation. The benefits of their complementary nature and specific features in evaluating the effects of projects are detailed below.

4.1.1 A quantitative evaluation system enhances qualitative analysis

The decision to use quantitative methods enhanced the qualitative evaluation at three levels.

- Firstly, through an analysis of the counterfactual and the use of statistics and econometric models, the application of quantitative measures allows **a precise measurement of a project's effects (or at least some of the effects) on the beneficiaries and a scientific demonstration of their causal relationship to the project**, which a qualitative analysis cannot do. Such an empirical demonstration of a project's effects is more convincing and rigorous than a factual argument of the achievement of the outcomes (qualitative analysis).
- Secondly, although the objective of the quantitative analyses was not to issue a conclusion on the achievement of a project's global impact, certain numerical findings on the outcomes can be extrapolated and sometimes generalised. This means that **some quantitative results can help to assess the global impact**, or at least help to support strong arguments for or against the achievement of a project's impact.
- Finally, through the primary data collection system established for the application of quantitative methods, **more and better quality qualitative data is collected** than if it had been collected as part of a standard qualitative analysis. Indeed, a quantitative method requires this data to be **collected randomly across a large number of beneficiaries and non-beneficiaries**, and sometimes at different levels for the same project (for example, at the level of households, villages, school principals, teachers, borehole managers, etc.).

The collection method in the field is also important: the **exploratory missions** are rich in qualitative data collection; the **household surveys** are packed with statistically exploitable qualitative information; and the **relatively long-term field presence** provides a large number of interesting observations¹⁸.

The fact that there is a greater variety of reliable qualitative data significantly increases the possibilities for assessing the global impact. Furthermore, all this qualitative data collected because of the establishment of the quantitative evaluation process helps to understand the mechanisms and context elements inhibiting or promoting the achievement of the outcomes and impact.

A data collection system for applying qualitative methods alone cannot provide this amount and level of detail in the qualitative data.

The three evaluations are good examples to illustrate how the quantitative system led to the realisation that it is possible to make qualitative analyses more rigorous. The quantitative system implemented and the time spent in the field have been effective in providing a rigorous argument on the achievement of the results.

The evaluation system was implemented as follows:

- (i) **Exploratory mission to prepare for data collection.** The existing project reports proved to be insufficient in order to design the questionnaires for the household survey and the interview guides for the focus group discussions. The evaluation team had to go into the field to

¹⁸ On average 4 weeks to carry out a household survey as part of the evaluations performed. This may be significantly longer depending on the project's scope and the study's objectives.

verify certain facts in the documents and to discover the current context through meetings with the beneficiaries and the local authorities.

All these exploratory missions proved to be extremely useful.

In **Morocco**, it led to the realisation that the list of beneficiary villages found in the reports contained many errors and that there was no list of beneficiary farmers in the villages selected by the project. These lists are obviously important when visiting villages to conduct a survey. These findings also highlighted a series of faults in the project.

In **Senegal**, the exploratory mission provided a better understanding of certain facts asserted in the reports that shed light on the achievement of the impact. For example in the reports, private standpipes appear as a project output. However, this is an outcome since they are financed by the users themselves through the distribution network constructed by the project. Secondly, also according to the reports, some standpipes were no longer used because of a lack of maintenance when in reality they are no longer used since they are no longer needed as the majority of users now have access to water in their compound.

In the **DRC**, the exploratory mission in Kinshasa led to an understanding of a series of adverse incentives in the institutional systems inhibiting any possible effect of the project. However, the decision was made to continue the evaluation exercise by visiting the province in order to triangulate the initial findings.

In **Rwanda**, such a mission has not been scheduled. It would probably have led to a faster realisation about the complexity of conducting the project evaluation in the current political context.

- (ii) **Primary data collection.** There is an important level of detail in the household questionnaires. These data collection tools can be used to ask numerous qualitative-type questions (assessment and open questions). The answers to these questions have been standardised, encoded and then analysed using statistical methods and/or used to formulate a rigorous argument.

The questionnaires used in the project evaluations in **Senegal** and **Morocco** illustrate the level of detail and the type of qualitative data collected and analysed (see the country reports in the appendices).

- (iii) **Time spent in the field** for data collection. Due to a relatively long field presence, the evaluators could spend more time on conducting interviews and observing and obtaining ad hoc information not contained in the interview guides or questionnaires. All of these elements help with the interpretation of certain findings.

In **Senegal**, a simple observation on the location of the public standpipes in the villages visited led to an understanding that the users of public standpipes travelled the same distance as before when they were going to the wells. Indeed, the well and the standpipe had been almost systematically located in the same place. Contrary to assertions made in several documents, no reduction in the distance travelled could therefore be expected. Certain unexpected effects were also discovered due to the evaluator's field presence. For example, the team accommodated in a health centre was able to observe that access to tap water increased the presence of civil servants (teachers, midwives) in the village.

In **Morocco**, the field presence enabled the evaluator to become aware of certain manipulations in the answers given by respondents, but also to be more critical of the relevance of the counterfactual.

In the **DRC**, the level of infrastructure development observed in schools has made it easy to understand why modern equipment had no chance of being used. Furthermore, the case of the DRC is special because no quantitative method has been applied given the revision of the methodological choices following the exploratory mission (cancellation of household surveys). However, the evaluation process presents certain quantitative aspects such as the random choice of schools, the use of a counterfactual, the formulation of hypotheses underpinning each outcome and the proposed tests for assessing the outcomes subject to the existence of data (and more fundamentally the existence of an impact).

4.1.2 A sound qualitative analysis is necessary for the rigorous implementation of quantitative methods

Quantitative methods truly enhance the evaluation of a project's effects as demonstrated above. However, to be effective, the application of such methods requires rigorous implementation which creates several challenges (see above for a detailed explanation): (i) finding a good counterfactual while minimising the problem of selection and contamination bias; (ii) using enough quality data; (iii) correctly applying the statistical and econometric techniques; and (iv) correctly interpreting the quantitative analysis results in line with the realities in the field.

A good qualitative analysis is required to tackle these many challenges that could ruin a quantitative analysis, especially in an ex post evaluation situation.

In each phase of the quantitative evaluation system, the qualitative analysis is used to justify methodological choices and interpret the quantitative analysis results.

The different phases of the *ex post* quantitative evaluation system are described below, where possible illustrating the required rigour and usefulness of the qualitative analysis in its implementation. Each of these steps is crucial and can be a source of errors affecting the quality of evaluations conducted using quantitative methods.

(1) Preparatory phase

- Nothing can begin without **a good understanding of the project's intervention logic, its objective and the context of the country, the intervention and even the context at the time of the evaluation.** All this is addressed by qualitative methods.
- For the application of quantitative methods, the next step is to **identify the measurable outcomes and the data type needed** to quantify them and demonstrate their attribution to the project.
- This means, at this stage, **precisely defining the relevant indicators** to perform the analyses of the desired outcomes. Note that it is vital to state precisely the definition of the indicators and specification of the information sources used to create them. Without precise data it becomes difficult to interpret the results.

The example that follows illustrates the importance of precision when defining an indicator. An indicator in the PARPEBA project in **Senegal** is the "coverage rate of access to drinking water".¹⁹ But what access is it talking about: private or public standpipe, at what minimum distance from the house? How is drinking water defined: chemical and/or bacterial composition? What water is it talking about: water from the water tower or from the tap or even at the point of use?

¹⁹ Furthermore, note in passing that the MDG definitions are extremely vague. It is therefore extremely difficult to issue a conclusion on a project's contribution to the achievement of the MDG.

The following example shows the importance of data sources (and their reliability). For the project in the **DRC**, there were several ways of measuring school attendance rates: through attendance registers (attendance rates will probably differ if attendance is taken in the morning or at the end of the day, in the rainy or dry season) or through a household survey or even through provincial registers (with data manipulation risks). The analyses will likely be different depending on which source is considered to build this indicator.

Angus Deaton stresses this in his contribution at the Conference on Impact Evaluation hosted by the AFD and EUDN in 2010 ("*Measure for Measure*"): "*different data leads to different conclusions*".

Those responsible for an evaluation must therefore argue that they have the appropriate data for their analysis by providing quality assurances.

One way is to present the detailed methodology and any bias that may result from the use of certain data.

- The preparatory phase is also the time to **design data collection tools** (questionnaires, interview guide) and **choose a credible counterfactual**.
- This is only possible with good field knowledge, gained in part through an **exploratory mission**. Following this mission and based on the qualitative observations made, arguments are put forward for the methodological choices with the aim of overcoming any identified bias (selection, contagion bias, relevance of the counterfactual, etc.). The survey tools are also improved.
- It is during the preparatory phase that the **local partner** must be contacted and a **field consultant** identified.

The first is needed for the logistics and implementation of primary data collection. He helps to build a team of reliable and competent enumerators, organises transport, gives advice on accommodation for the team, etc. Ideally, this partner should be knowledgeable about quantitative evaluation but also be aware of realities in the field. Research centres or universities can be good partners.

The second is a person close to the project in the field without a conflict of interest. It is important that this role is clearly defined and their presence has no effect on the respondents' attitude. Being accompanied in the field can help to ease introductions with the local population and gain their trust by overcoming certain cultural barriers.

Without such support, the implementation of a rigorous primary data collection system is almost impossible.

- This exploratory mission is also an opportunity to **obtain the approval of the country's national authorities** to conduct this type of study.

In **Rwanda**, the Ministry of Health expressed strong reservations about the evaluation of the project selected, perhaps out of fear that the findings could have political consequences. Although the team and the OSE tried to earn the trust of the authorities and obtain their approval to carry out secondary data collection, not all permissions were obtained. The evaluation therefore had to be stopped²⁰.

²⁰ Unfortunately, all this was discovered during the field mission, because there was no exploratory mission planned due to budgetary constraints.

(2) The survey phase

- Firstly, this involves **testing the data collection tools** to guarantee the proper formulation and a good understanding of the questions, particularly in the local language.
- The **competences and reliability of the enumerators** must also be verified.
- It is essential to **adapt constantly to the realities in the field without losing sight of the rigour needed** to be able to apply quantitative methods to the data collected. There are many hazards in the field, a flat tire, adverse weather conditions, the absence or unavailability of respondents, impractical lines of communication, etc. The partial review of sample size or type is common.

Be aware that a significant reduction in the sample or changing the geographic area covered could challenge the feasibility of using quantitative methods (use of econometric and statistical tools) and therefore cast doubt on the credibility of the analyses.

It is therefore a question of being careful not to introduce significant biases that may harm the quality of the assessment of the outcomes. There are several actions that are useful in minimising these biases: collecting as much information as possible that can **demonstrate the relevance of the counterfactual** (qualitative observations and quantitative indicators), as well as the **relevance and reliability of the data collected**. It is important to take the time to earn the trust of respondents in order to avoid manipulated answers leading to over- or under-estimates of the effects. It is vital that the questionnaires are re-read at the end of each day and the enumerators asked for explanations in the event of inconsistencies or no answers. If necessary, the enumerator must be sent back to the respondents to ask them further questions.

- Finally, the step that brings the field phase to an end consists in obtaining a usable database. This means creating an **input mask** in a computer programme (CSPRO for example) and finding a team to **encode the data collected on the questionnaires**. This must be supervised by an experienced individual. Indeed, good data on paper is no guarantee of good encoded data and poorly encoded data results in flawed analyses.

(3) The analysis phase

- Firstly, the data must be **cleaned** to verify that the encoded information is consistent with that collected on paper.
- The appropriate quantitative methods for the type of available data must then be chosen and any bias identified (selection problem, reliability and relevance of baseline data, validity of the counterfactual, etc.).
- In addition to identifying methods, they must be **used correctly** (respect the hypotheses). STATA, a statistical and econometrics software package was used for the quantitative analysis.
- Then, it is important to ensure the **robustness of the results** (identical results across different models and with different specifications).
- Finally, the results must be **interpreted in light of the field observations**. It is essential to remain critical of the figures put forward by the methods, especially if they contradict intuitions in the field. In this case, there must be a check that there is no technical manipulation error. This final stage in the qualitative analysis is crucially important.

4.1.3 Qualitative analysis is the key to understanding a project's effects (outcomes/impact) and assessing the impact

The intervention logic and context are two key elements to understanding an intervention and its impact. In addition to being useful in justifying the establishment of a quantitative method, deciding between the methodological choices and interpreting the findings on outcomes, analysing the intervention logic and its context are essential to supporting a reasoned judgement on the achievement of the effect at a general level (impact). Indeed, quantitative methods are limited to assessing outcomes. However, empirical findings on the outcomes can be used to argue for a qualitative assessment of the impact (see above).

Highlighting the **causal sequence** between the resources mobilised and the objectives to be achieved firstly provides a critical overview of the intervention's feasibility. Are the actions likely to lead to the objectives? Are those objectives realistic in view of the context and the resources mobilised? What are the risks and underlying hypotheses? All these questions help in understanding the intervention from its design to the impact.

Knowledge of the intervention's context is a determining factor in understanding its impact. The socio-economic, political and cultural situation at the time of the intervention, diplomatic relations between the countries, the relationships between donors, comprehending the leadership challenges at different levels of power and the historical perspective of the intervention²¹ are just some of the contextual elements that may influence whether or not an intervention's impact is achieved.

For the three evaluations, **these aspects have proved essential to the analysis and have helped to show the shortcomings in the project design and implementation or monitoring which prevent or reduce the expected impacts** (see below).

4.1.4 A system that allows all actors to be considered

Combining two methodological approaches allows for **in-depth interviews with beneficiaries through the quantitative approach and a more precise assessment of the roles of the actors involved in an intervention through the qualitative approach.**

Indeed, the use of qualitative methods requires information to be collected from the BTC, the DGD, the partner (ministry responsible for the sector, local authorities) and several beneficiaries. This information is very useful for, among other things, putting the intervention into context. As for quantitative evaluation, this prioritises the large number of beneficiaries chosen at random. The combination of these two methods therefore enables more in-depth interviews to be conducted with the various actors involved in an intervention's success. The benefit of this is that it takes into account the specific features of each actor and raises their awareness of impact evaluation.

4.1.5 The feasibility and relevance of an ex post quantitative evaluation

Given the demand and rigour necessary to conduct evaluations using quantitative methods, as well as the time and costs that it represents, the issue of the feasibility of a quantitative evaluation was raised for the 4 projects to be evaluated. A good qualitative analysis of the intervention logic and context at the time of the project and during the

²¹ For example: are there other interventions of this type? By the same donor? By other donors? Have they been well or poorly received by the population? Are other interventions of this type planned?

evaluation, complemented by an exploratory mission has allowed a conclusion on the relevance of implementing a quantitative evaluation to be determined.

According to Naudet et al. (2012), the experimental method is suitable for projects that they characterise as a "tunnel programs", i.e. interventions characterised by a limited number of homogeneous inputs, a clear and tested causal sequence, broad participation and rapid effects. This experimental analysis is only partially relevant in the case of the four evaluations conducted, because they are ex post evaluations with a quasi-experimental design. Despite conditions that were far from optimum for performing quantitative analyses, the initial methodology proposed for the four projects systematically included a quantitative dimension. However, ultimately, only two evaluations (Senegal, Morocco) out of the four use quantitative methods.

The situations of the four projects in terms of the feasibility of a quantitative analysis are presented below. Conclusions are then drawn from these experiences. Finally, the question of the relevance of conducting quantitative evaluations although others have been conducted on similar themes is addressed.

1) The feasibility of a quantitative evaluation in the 4 evaluated projects

a) Supporting technical and vocational education in the DRC:

- **Facts:** this project has a quasi-national scope with a certain sequential implementation since the previous project, which was relatively similar, concerned other provinces. It has not been possible to use secondary data for two reasons: no recent, good quality data has been collected in the provinces and in particular the low probability of seeing the project's effects at province level given the very small number of schools affected by the project in each province. No data was available at project level (for phase 1 or phase 2) and it was impossible to visit all the provinces to collect information. Furthermore, the use of a telephone or email collection system could not be considered due to the underdevelopment of this infrastructure in the country.
- **Initial proposal:** the initial methodological note proposed a case study with counterfactual in a province chosen based on objective criteria with surveys with the different types of beneficiaries (schools, teachers, students, graduates).
- **Findings after the exploratory mission:** this mission showed that there was a high probability of not finding an effect of the project on the beneficiaries. The establishment of a primary data collection system through individual surveys has therefore been revised.
- **Analyses performed:** the methodological approach remained the same with individual surveys being replaced with focus groups with the same groups of beneficiaries and a counterfactual. In-depth interviews have been conducted with those involved in the sector but also with school principals, teachers, students and graduates. Secondary data at school level has also been collected. All the data collected has allowed a comparison of the relevant indicators for each outcome between the target schools and the control schools for 4 sectors of interest. These comparisons, combined with an analysis of the institutional context, have enabled a rigorous qualitative analysis on the non-achievement of the desired effects of the project on both a beneficiary and general level.

b) **Supporting the construction of a public health reference laboratory project in Rwanda:**

- **Facts**: this project has a national scope with sequential implementation of quality control. There is relatively good quality secondary data in the districts and the laboratories.
- **Initial proposal**: it was proposed that this secondary data be used through sequential implementation to create a valid counterfactual. In-depth interviews with a sample of laboratory heads, health centres and hospitals were also planned.
- **Findings during the field mission** (there was no exploratory mission): data was indeed available to study at least one aspect of the project (the control of thick smear analyses). However, since the team never received official authorisation to collect this data and therefore analyse it, they were unable to complete the evaluation (even though the qualitative part was conducted during the mission).

c) **Project to supply drinking water and improve sanitation in Senegal:**

- **Facts**: this project focuses on 3 regions in the Arachidier Basin without sequential implementation. The project comprises different components and dimensions: access to water through the rehabilitation or construction of boreholes, water purification techniques to attempt to reduce fluorine levels at two pilot sites, construction of sanitation facilities and hygiene awareness. There are participatory diagnostics containing quantitative and qualitative information about the situation on certain sites prior to the intervention. However, the raw data is no longer available and therefore cannot be used for the ex post impact evaluation. There was no predefined counterfactual for the project.
- **Initial proposal**: the methodological note proposed conducting a household survey with counterfactual in one of the regions affected by the project, concentrating on the effects of access to water via the new boreholes. The sanitation and water purification components were not subject to a quantitative methodological design due to a lack of time and resources. Focus groups in the villages and with the borehole managers were proposed in addition to household surveys. Reconstruction of baseline data was planned using recall data and through the information available in the participatory diagnostics.
- **Findings after the exploratory mission**: it was possible to find a credible, although imperfect counterfactual in the survey area with a small sample. There may have been overestimation bias of the effects since the control group was less well served with water than the target group, even before the intervention. Furthermore, there was a spill-over effect: households living in villages not connected to a borehole were also being supplied with tap water. Data on the pre-intervention situation was able to be collected (recall), but not for all the outcome indicators. Consequently, this limited the use of quantitative methods. It was also observed that the participation of beneficiaries was almost unanimous. Thus, selection bias was highly unlikely and as a result the small sample size was less of a problem. The identified local partner was of a good standard and had already assembled a team of enumerators. The field consultant found was well-suited for the mission. The reception in the villages was warm and the respondents made themselves easily available.
- **Quantitative analyses performed**: the data collected allowed the use of quantitative methods on two outcomes (effect of the project on the

distance travelled to fetch water and the time allocated to this task). Other outcomes were subject to statistical analyses but were unable to demonstrate their attribution to the project (effect on education or well-being for example). The data collected also helped to enhance the qualitative analysis on several levels.

d) Project for small and medium hydraulics in Morocco:

- Facts: this project is concentrated across three municipalities on the Tiznit Plain. It focuses primarily on the rehabilitation of traditional flood irrigation systems in the villages situated along the main river. It also comprises other components: reforestation and capacity building on both cultivation techniques and the management of rehabilitated structures. The project has no baseline data or predefined counterfactual.
- Initial proposal: the plan was to conduct a household survey with counterfactual covering the 3 municipalities affected by the project and another 3 "control" municipalities situated along a tributary of the main river.
- Findings after the exploratory mission: there was no data, even at the level of the project's beneficiary villages. Creating the sample was not easy. Contact with the local authorities proved to be difficult. They tried to influence certain methodological choices and the answers of some of the people questioned. The local partner was not experienced in this type of individual survey. It was also found that it would be impossible to collect quantitative data on the land yields with the available time and budget.
- Quantitative analyses performed: quantitative methods were successfully applied for two outcomes (irrigated areas in high and low flood periods; the farming income share of total income). However, it is important to remain cautious in the interpretation of these results which could be overestimated given the doubts raised regarding the credibility of the counterfactual on certain aspects. In addition, since matching was done on ex post data and because some results are not robust across various specifications, the report is cautious regarding the interpretation of the results found.

2) Conclusion on the feasibility of quantitative analyses in the evaluation

From the experiences, **four facts** emerged influencing the feasibility of a quantitative analysis for the effects of an intervention.

- a) It is essential to have a credible counterfactual to be able to conduct quantitative evaluations of the outcomes, in the sense of "attribution analysis" (3ie).

The use of quantitative tools is strongly contested

- when the project has a national scope without sequential implementation of the intervention (known as the pipeline process); or
 - when problems of contamination or selection bias cannot be eliminated or at least minimised by a rigorous qualitative argument and/or solid econometric techniques.
- b) It is necessary to have or be able to reconstruct baseline data. The **baseline** is another important element in conducting attribution analysis. Without a baseline it is not impossible to use certain quantitative tools, but their application is more difficult to justify (see above, matching on ex post data).

- c) The feasibility of quantitative impact analyses also depends on the **agreement of the local authorities**. Indeed, the advantage of rigorous quantitative analyses is that they provide evidence of the existence (or the absence) of effects related to the project. These scientifically proven findings may therefore have more substantial political implications because they are less easily refuted than a reasoned judgement on the achievement of a project's effects as proposed by a qualitative analysis.
- d) The **value for money** of implementing a quantitative evaluation process. If a rigorous qualitative analysis can be used to show the probability of observing a very weak effect, there is no need to collect primary data. Indeed, the quantitative analyses of this data would not contribute anything more. Furthermore, a very large sample would be needed to justify that the lack of effect was not due to the sample size. The opposite is not true: if a qualitative analysis shows a reasonable probability of observing an effect, then it is useful to implement a quantitative analysis process. In fact, this would allow a conclusion on the extent of the effects and demonstrate empirically the causal relationship with the project. This is more rigorous and convincing than any qualitative analysis.

For example, in the **DRC**, the decision was made to abandon the household surveys having shown that the ATEP2 project in the specific context of technical and vocation education in the country had a high probability of not achieving the desired outcomes and thus the expected impact. The efforts needed to carry out a quantitative analysis were therefore no longer justified.

It should be noted that the issue of value for money concerns a much broader aspect of impact evaluations. The cost/benefit analysis will never be positive if the benefits of a quantitative evaluation are not perceived. This perception refers to the use of an evaluation's findings. In other words, why would evaluations be needed if not to use their findings? The expense of implementing a quantitative evaluation system can never be deemed acceptable if the measurement of the effects on the beneficiaries and evidence of their attribution to the project is of no interest to policy makers and/or development practitioners.

Three observations deserve further attention. They concern the scope of quantitative analyses.

- a) Quantitative methods can be used to assess **one aspect of a project** (one particular component or even one dimension of a component), but it is more difficult for them to evaluate all the activities of an intervention. Easily identifiable and measurable effects are preferred. Other project aspects and outcomes that are difficult to measure can then be analysed qualitatively.
- b) Even if the project has a national scope or concerns several regions, a quantitative analysis of the outcomes can be performed on a **limited geographical area**. Analysis of a province, municipality or region has already provided many lessons. A good knowledge of the realities in other regions sometimes allows certain findings to be extrapolated or even generalised across all the areas affected by the project.
- c) **Quality secondary data** is sometimes sufficient for conducting quantitative analyses, especially if it is supported by a rigorous qualitative analysis. Ideally, this data would have been generated by the project's monitoring-evaluation process (see next point).

The exercise conducted as part of these four ex post impact evaluations shows that even in a sub-optimal situation, it is possible and appropriate to implement an evaluation

system using quantitative methods if certain conditions are met.²² These findings led to the creation of table 2 below.

This table summarises the conditions to be met in order to ensure the **feasibility of rigorous quantitative evaluations of quasi-experimental design** (as opposed to the experimental design of the randomized control trial which is unrealistic in Belgian bilateral cooperation) **in an ex ante and ex post situation**.

Depending on whether the outcomes and impact evaluation is planned when the project is designed (ex ante evaluation) or at the end of a project (ex post evaluation), certain conditions must be met to ensure the feasibility and relevance of a quantitative evaluation of a project. The objective is to quantify the changes observed on the beneficiaries following the intervention and demonstrate empirically the causal relationship.

In a quasi-experimental design, the minimum and optimum conditions influence the quality level of the potential quantitative analyses.

In addition to an expertise in the tools, the quality of a quantitative analysis depends on the following elements:

- The **relevance of the identified counterfactual**. It must be proven that it is valid through statistical analyses but also by a reasoned judgement based on the qualitative analysis of the context and field observations.
 - o If a project experiences excessive contamination effects on the non-beneficiary population, the application of quantitative methods will be seriously challenged.
 - o The sequential implementation of an intervention (pipeline process) is one satisfactory option in terms of methodology which is not too restrictive for the project's implementation.
- The **quantity of data** collected.
- The **reliability and relevance of the data** to be analysed. Recall data is still judged less reliable than factual data at the time of the surveys. The quality of secondary data sources used must also be justified.
- The **ability to minimise or at least identify elements that could bias the results** (principally contagion and selection bias) through econometric models and arguments taken from field observations and interviews.

It is worth mentioning two elements:

- The **baseline**, even done well, is a necessary but insufficient condition for the application of quantitative evaluation methods.
- The decision to design an **ex ante impact evaluation** offers more opportunities to be able to conduct quantitative evaluations. Furthermore, an ex ante design may lead to a change in a project's implementation. For example, a sequential implementation may be favoured in order to create a relevant counterfactual (**pipeline process**).

The required conditions for a successful quantitative evaluation also appear in the table. If these conditions are not met, there is no use implementing a quantitative evaluation system. Finally, the three observations regarding the scope of quantitative evaluations are summarised in the final row of the table, entitled "alternatives". These alternatives can be used to limit the scope of the quantitative evaluation and therefore inevitably its cost, but with consequences on the number of attributable effects that can be analysed.

²² For a more in-depth discussion on how to conduct evaluations under tight time and budget constraints, with limited access to data, the following article is recommended: "Conducting quality impact evaluations under budget, time and data constraints" -Impact Evaluation Group (IEG) of the World Bank.

It must be stressed that, even if the optimum conditions are met, it is **impossible to overlook a rigorous qualitative analysis**. This is essential for arguing for certain methodological choices and interpreting quantitative results. **It is not enough to demonstrate what works or what does not work, we must also explain why and how we have arrived at such findings.**

Table 2: Conditions for conducting quasi-experimental, ex ante or ex post design quantitative evaluations.

Quasi-experimental design	Ex ante evaluation design	Ex post evaluation design
Required conditions	<ul style="list-style-type: none"> - Approval of local authorities - Partner collaboration - Incentives to implement a results based M&E system (human resources, budget, skills) 	<ul style="list-style-type: none"> - Approval of local authorities; - Relatively high probability of finding an impact - Enough time and a sufficient budget - Plan an exploratory mission
Minimum conditions	<ul style="list-style-type: none"> - Identify a credible counterfactual (without excessive contamination) - Enough reliable and relevant baseline data on a sample of beneficiaries - M&E system allowing results based management (outcome/impact) → monitoring data - Enough reliable and relevant ex post data on the beneficiaries and counterfactual - Reconstructed baseline situation for the counterfactual based on recall data and/or quality secondary data 	<ul style="list-style-type: none"> - Identify a credible counterfactual (without excessive contamination) - Enough reliable and relevant ex post data on the beneficiaries and counterfactual - Reconstructed baseline situation for the beneficiaries and the counterfactual based on recall data and/or quality secondary data
Optimum conditions	<ul style="list-style-type: none"> - Identify a credible counterfactual (without excessive contamination), for example through the sequential implementation of the intervention (pipeline process) - Enough reliable and relevant baseline data on a sample of beneficiaries - During the project's implementation, identify a quality secondary data source to reconstruct the baseline of the counterfactual - M&E system allowing results based management (outcome/impact) → monitoring data - Enough reliable and relevant ex post data on the beneficiaries and counterfactual - Collect recall data to complete the baseline of the counterfactual 	<ul style="list-style-type: none"> - Identify a credible counterfactual (without excessive contamination), if necessary using the sequential implementation of the project - Enough reliable and relevant ex post data on the beneficiaries and counterfactual - Reconstructed baseline situation for the beneficiaries and the counterfactual based on recall data and/or quality secondary data (through the M&E system that collected the relevant indicators - requires access to raw data).
Alternatives	<ul style="list-style-type: none"> - Focus the quantitative evaluation on one aspect of the project or one geographical area - Depending on the objectives and type of intervention, reduce the number of individual surveys to concentrate on a higher level of beneficiaries (villages, municipalities, schools, hospitals, health centres, etc.) 	

3) Reasons for conducting quantitative evaluations when evidence exists

One of the evaluation's objectives is to capitalise on experiences, i.e. to conduct a large number of studies showing what works and what does not work in terms of development assistance. Some people could therefore be tempted not to try and assess the impact of their intervention, but to reproduce their intervention based on

a successful project. In such a case, they would only need to focus on monitoring and not evaluation.

It is certainly interesting and important to draw conclusions from existing quantitative evaluations when designing a project. However, starting from the hypothesis that the implemented project will produce the desired effects because it is based on a similar project that has already been positively evaluated is risky for several reasons.

- a) **The limited external validity of randomized control trials.** RCTs are the quantitative evaluations most able to provide evidence of an intervention's success. However, RCTs are often conducted on a small scale and largely used by academic researchers (optimum conditions sought for applying quantitative methods). It cannot therefore be ruled out that the results will be relatively different, once the project is developed and implemented on a larger scale by a development agency with local partners and without academic support. (White, 2013, 3ie).
- b) **The importance of the intervention mechanism.** RCTs do not always seek to evaluate the intervention mechanism, but instead attempt to answer a research question affecting one or another aspect of the project. However, the intervention mechanism is a determining factor in development assistance.

Studies increasingly evaluate the intervention mechanism.²³ However, the implementation of such mechanisms, developed in a research perspective is not always obvious as real-world conditions influence the implementation process.

- c) **The importance of the context and its rapid development.** It is one thing to copy interventions that have proved successful, but another to be able to reproduce the conditions that influenced its success. Indeed, all the rigorous evaluations performed focus on the importance of the context elements that promote or inhibit the achievement of a project's objectives.

Furthermore, it is rare to find evaluations that focus on both the themes and geographical area of interest. In fact, the context in the broadest sense must be considered as a determining element in an intervention's success: the social, political, economic and cultural aspects, along with the historical dimension.

In addition, the fast-changing society in the Southern countries (telecommunications, climate shocks, etc.) and thus the changes in individual behaviours must also be taken into account.

²³ For example, there are several possible mechanisms for promoting education, among other things, paying money to parents to encourage them to send their children to school (cash transfers) or a programme guaranteeing a school meal (food for school) or even the payment of school fees through certain sanctions if the child has a high rate of absence. Only a rigorous quantitative analysis of the outcomes can help decide between the effects of these different modes of intervention.

4.2 Lessons on the projects' design, implementation and monitoring-evaluation

The outcome and impact evaluations have highlighted shortcomings in the projects' design, implementation and M&E system.

There is no denying that for the four evaluated projects, **the assessment of the achievement of the outcomes and global impact is not good** if it is based on the specific and global objectives expected for each project. However, if referring to the specific and global objectives that the evaluators redefined following the reconstruction of the logical framework, the assessment is slightly better, although far from good.

This indicates two things. Firstly, there are shortcomings in the projects' design/formulation: the objectives are not in line with the realities in the field and the intervention's activities/resources. Secondly the projects' implementation and M&E system is not effective enough to ensure results based management (outcomes and impact).

Since then, some improvements have obviously been made in these areas, because these findings concern projects completed after 2008. However, although practices have changed, it would be advisable to verify whether they mitigate the shortcomings identified below. It should then be checked whether these changes are truly reflected in the field and in the observed results (outcomes and impact).

4.2.1 Shortcomings in the interventions' design/formulation which reduces their potential impact

In the four evaluated projects, several issues have been identified in their design/formulation.

The first problem is that all the projects have a **weak intervention logic** reflected by:

- **Objectives disconnected from the activities and resources** used.

Links between the activities and the specific and global objectives do not always exist or are not clearly established in the intervention logic. Consequently, even if all the activities are adequately performed, the effects on the beneficiaries may not occur and these activities may not contribute to the achievement of the global objective.

In the **DRC**, the equipment was properly delivered, but most of it is unused. In **Senegal**, the taps are connected and people therefore have access to water, but the quality is not guaranteed, even though the objective was access to drinking water.

- **Unrealistic objectives** in the local context.

The projects often have unrealistic global objectives given the available resources. For example, this is the case in the **Congo** where the project providing technical support to schools aimed to improve the employment situation in the Congo, although it targeted barely 5% of professional schools with just a few beneficiary schools per province.

All these projects are characterised by a **lack of knowledge about the context and the problems to be resolved**, even though 3 of the 4 projects are in fact the second phase of a previous project.

For example, the project objective in the **Congo** was to increase the availability of a qualified technical workforce even though there are no jobs in the targeted sectors and the aim of almost all the young people who enrol on these courses is to continue their studies in higher education and not to go into manual work. For the same project, most of the equipment delivered requires an electric power supply, although this is lacking even in Kinshasa. In **Morocco**, the project's objective was to reduce the rural exodus, although this region can only survive through income diversification since flooding and therefore farming during floods is random, even with one of the rehabilitated irrigation systems. So working in the city is still an attractive alternative for these people. In **Senegal**, the project's objective was to supply drinking water, although the region has serious problems with ground water potability and 2 out of 3 boreholes contain water that is too high in fluorine and sodium. Still in Senegal, one of the proposed water purification techniques was the use of calcined bone, although for around fifteen years the literature has demonstrated this method to be ineffective.

- **Vague and poorly defined objectives** (no target value and/or clear definition)

The objectives systematically lack precision in their definition. Few objectives are quantifiable, and how the effects on the beneficiaries are to be measured is not clearly specified. Furthermore, given that there are no baselines, even if certain objectives can be quantified, it is impossible to assess any changes.

For example, the objective of the project in **Senegal** was to increase the rate of access to drinking water in the region by 50%, but how can any change be assessed without a pre-project indicator? What is meant by access to water, at home or at what maximum distance from the home? What is meant by drinking water, the chemical and/or bacterial composition?

Another problem is the **issue of alignment with national policies**. According to the commitments on aid effectiveness (Paris Declaration, Accra, Busan), this is an important, if not essential condition for achieving impact. But alignment is not a guarantee of a project's success. Indeed, certain elements of national policies may be contradictory to the intervention's success as defined in the specific and global objectives.

For example, when the PARPEBA project in **Senegal** was designed, the government supported a policy of access to water for as many people as possible, with no concern for the quality. The project was aligned with this policy, which led to the construction or rehabilitation of most boreholes with harmful fluorine and salt levels, even though it wanted to supply drinking water. In **Morocco**, there was no participatory aspect included in the rural development policies which strictly prioritised physical achievements. The PMH-Tiznit project was part of this policy framework although it has long been demonstrated that an irrigation project has a greater chance of sustainable success when the beneficiaries are actively involved in its design and implementation.

More fundamentally, these findings on alignment raise the question of the **ability to negotiate with the partner country** during a project's design. They also raise the question of the **ability to abandon or suspend an intervention** when Belgium believes that the conditions for achieving the desired outcomes and expected impact are not met.

A weak intervention logic and a "blind" alignment (i.e. an alignment with national policies that goes against a project's success) **significantly reduce the chances of an intervention's success**.

These weaknesses observed in the projects' design is concerning since they show that **the impact is not a priority and that the logical framework is a formality for presenting a project even though it could be a real tool for formulation, results based management and evaluation**. Indeed, a simple examination of the logical frameworks would have uncovered the impossibility of achieving most of the expected outcomes before the evaluated projects began.

In part, these findings can perhaps be explained by the method of identifying and formulating projects even within direct bilateral cooperation. The DGD holds political dialogues with the authorities to establish the specific and global objectives of the interventions during the identification phase. Then, the BTC formulates a project with the aim of achieving these objectives. This means that the inputs and outputs are defined by the BTC, while the outcomes and impact are defined by the DGD. **Although there is a consultation process between these two bodies and the partner country, it is not working because it is allowing unrealistic projects.**

4.2.2 Implementation that favours disbursement and physical achievements

For the four projects, the **expected outputs have effectively been achieved**. This is a positive finding that deserves to be highlighted insofar as the interventions examined were often deployed in difficult conditions in terms of the context or the relationship with the national authorities in question. An examination of the monitoring-evaluation documents reveals the efforts that had to be made to achieve these results.

However, what about the effects of the achievement of these outputs? In other words, are these outputs being used? Are they producing the expected effects?

It is clear that this **output focused management prioritises disbursements and physical achievements to the detriment of global and specific objectives**.

Without results based management, in the sense of outcomes and impact, a project's effects become secondary objectives. This is clear from the evaluated projects, for example in terms of the decisions made in response to unexpected and/or unanticipated situations. In **Senegal**, when the project was faced with the problem of a significant number of boreholes with poor water quality, the decision was made to ignore standards and connect these boreholes. Yet, such a decision deviated from the project's objectives. In **Morocco**, given the topography of the locations, more detailed technical studies should have been conducted in a certain number of areas. This would probably have improved the ability to irrigate in flood periods. In the **DRC**, although the schools had not conducted the rehabilitations as agreed, they still received the planned equipment.

The implementation method observed only **reflects the lack of incentives to do otherwise**. If executing agencies are assessed in terms of disbursement and achievement of outputs, why would they seek to achieve the desired outcomes and expected impact?

4.2.3 Inadequate M&E system

Another illustration of this lack of concern for outcomes and impact is the lack of attention for these aspects in the project M&E systems. While, relatively precise recommendations on the actions to be carried out with a view to assessing the project's effects were formulated for all the projects, these were never implemented.

The three evaluations highlight the shortcomings within the **projects' monitoring systems**. Although the relevant indicators for measuring the outcomes had been defined in the project formulation documents, they were often poorly defined, without an initial value and/or target value or an explanation about the collection arrangements (how, when, where, by whom?). Finally, several years after the projects have ended, these indicators still cannot be found, even in databases (for example in **Senegal**, the raw data normally used to draft the participatory diagnostics is unavailable). This finding raises the problem of access to reliable monitoring data in order to conduct evaluations on the achievement of the projects' results (outcomes and impact). Furthermore, such data would greatly reduce the cost of collecting information during impact evaluations at the end of the project.

The **internal evaluation system** is also challenged. Indeed, almost all the conclusions issued by the final project evaluations have been contradicted or slightly modified by the findings of the ex post evaluations. In addition, it is unsurprising to note that the final evaluations are strongly focused on the achievement of outputs, in view of the priority given to disbursement and physical achievements.

Once again these findings raise the **issue of incentives to do otherwise**. Knowing that one euro invested in the M&E system is not invested in other activities, how can the partners and executing agencies be encouraged to dedicate more resources and means to monitoring and evaluation? Especially when they suffer no consequences for not doing so and do not perceive the added value of management that considers the results (outcomes and impact) and not just the achievement of outputs. Furthermore, the extent of the work to implement an intervention must not be ignored. It is not easy to juggle the implementation of a project and a rigorous M&E system.

5. Recommendations

5.1 Define a global evaluation strategy where impact evaluation has a place

Define a global evaluation strategy within Belgian development cooperation where impact evaluation has a place.

Finding: there is no global evaluation policy or strategy for all Belgian development cooperation, although evaluation and reporting on the results are provided for in the law of March 2013 on international cooperation.²⁴ Furthermore, if we look at what is practised within the bilateral cooperation of the neighbouring countries or international organisations, Belgium is clearly behind in the field of impact evaluation.

A global evaluation strategy is essential given the complexity of Belgian development cooperation interventions and the number and specific features of the various stakeholders. Since no such policy currently exists, situating impact evaluation (a subset of evaluation) is bound to be an incomplete exercise.

When talking about impact evaluation, throughout this document reference is made to **ex post project evaluations** that aim, through a combination of qualitative and quantitative methods, to measure the outcomes attributable to the intervention and formulate a rigorous and reasoned judgement on the impact of a development operation. Outcomes are defined as the effects on beneficiaries while impact corresponds to the general effects of the intervention. These terms correspond to the concepts of specific objectives and a global objective, as presented in the intervention logic.

Evaluation in general, but also outcome and impact evaluation, has three objectives: **(1) accountability; (2) decision support; and (3) capitalising on experiences**. The findings of an evaluation and especially an impact evaluation may interest the various cooperation partners (the donor and policy makers, partner, beneficiaries, executing agencies and taxpayers) who may use them depending on their needs. Indeed, impact evaluation consists in identifying "what works and what does not work" through an empirical demonstration, but also in understanding why and how a project does or does not achieve its objectives.

²⁴ "Art. 32. With a view to achieving Belgian Cooperation Development goals, such as those defined in chapter 2, the **results are evaluated** in light of the principles laid out in chapter 3 and on the basis of the criteria defined by the OECD's DAC, namely relevance, effectiveness, efficiency, feasibility, and **impact**, as well as on the basis of sustainability. To that end, a coherent approach will be drawn up with a view to enabling the **reporting of results and results based management**. A standardised reporting system should also enable the **systematic monitoring of the results obtained or not**. Furthermore, account will be taken of these results when deciding the future of the actions undertaken."

The questions about aid effectiveness and the types of projects to be implemented in order to generate an impact and have an effect on a country's development are central to the debates. In this context, there is a demand for **accountability** for the achievement of outcomes and impact of development assistance interventions. This involves accountability to taxpayers but also to partners and beneficiaries.

- The impact evaluation can also **help with decision making**. It can help to understand the opportunity to replicate an intervention or avoid doing so. An evaluation's results can therefore assist in making strategic decisions about the direction of aid in a sector, region or country.

Although this is theoretically possible and highly desirable, in practice the timing is often not favourable for the impact evaluation to play such a role. In this regard, the example of the **DRC** is an interesting one. While the decision to continue with a project in the technical and vocational education sector has already been made, the evaluation shows that support in this field has had no impact and there is a very low probability today that such support would have any effect on the beneficiaries in the current context.

Furthermore, impact evaluation can help to **provide a better basis for cooperation policies using evidence of their effectiveness** (evidence-based policy making). This could strengthen the position of the Minister for Development Cooperation during budget negotiations for example. Policy-makers must however be ready to accept the fact that certain rigorously demonstrated findings may challenge the relevance of certain actions or intervention methods in some sectors or countries. But are we not seeking more effective public spending for greater effects on the beneficiaries?

- **Capitalising on experiences** is a particularly interesting aspect of impact evaluation. This is an investment in the creation of a public good. Impact evaluation is used to identify those projects that work and those that do not, but it can also provide such a judgement on an intervention method (how a project is designed/implemented) or a particular aspect of a project. Even if the aim is not to replicate the intervention elsewhere or on a larger scale due to a lack of resources or political will, the results of a rigorously conducted evaluation can be published. They can then be used to argue the case with other donors to encourage them to finance (scale-up), or conversely, force them to abandon these types of intervention.

The project in **Morocco** readily lends itself to learning because flood irrigation is still relatively unknown. Likewise, the water purification component in **Senegal** could have been the subject of an evaluation with the aim of learning how large the problem of safe drinking water is in this country.

Furthermore, capitalisation can also be conducted on a formative or methodological level; as is the case with these four ex post impact evaluations.

Promoting impact evaluations and incorporating them into a general evaluation strategy must be done while taking account of several elements described below.

5.1.1 Clarify the definitions of impact, outcome and impact evaluation

A first key element is to **clarify what is meant by "impact evaluation", and what the terms "impact" and "outcome" mean**. It is also important to formulate what we want to know and what the **objectives** are when conducting an impact evaluation.

Indeed, one of the findings of this exercise is the plethora of definitions and misunderstandings generated when talking about impact evaluation. Between the DGD, the BTC, the OSE and others involved in development cooperation, it would be useful to have a precise and standardised definition of impact and outcome. The definitions presented in this report could easily be used as a basis for this discussion (see above, section 3).

It should be noted that it would be worth debating this issue on the definitions of impact at DAC level. This study is in line with a broad consensus (White 2010 and 20136, World Bank, 3ie) to challenge the current definition proposed by the DAC which defines impact as a long-term effect. It would also be worth standardising the definition of impact evaluation, clearly distinguishing between "impact evaluation" and "outcome evaluation" although these two elements can be evaluated within the same study as the exercise demonstrates.

5.1.2 Raise awareness about impact evaluation

The second element is to **raise awareness about outcome and impact evaluations**. Development cooperation practitioners seem to be relatively familiar with impact evaluation, but their knowledge remains relatively theoretical or abstract. This is because in the field, the M&E system cannot be used to assess the achievement of outcomes or impact.

These evaluations have shown that the achievement of outcomes and impact is still too often considered as a secondary objective compared with the achievement of outputs. In other words, priority is given to output-focused management rather than management focused on the achievement of the effects on the beneficiaries and at a general level (outcomes and impact).

In this context, outcome and impact evaluation is often overlooked because little is known about either the quantitative evaluation process or its interest for the various stakeholders in an intervention. It is worth highlighting the ability of impact evaluations to produce a wealth of useful information in order to convince stakeholders to use them into practice.

A quantitative outcome evaluation provides an objective and rigorous measurement of the effects on beneficiaries by demonstrating empirically their causal relationship with the intervention. A rigorous qualitative analysis provides an understanding of the factors and mechanisms that have inhibited or promoted the achievement of the effects. It also supports a reasoned judgement on the achievement of the impact. Such results enable policies to be implemented based on the evidence of their effectiveness. **The evaluation findings can be solid arguments, strengthening the position of a Minister of Development Cooperation for one case or another.** Promoting and raising awareness of impact evaluation means encouraging strong and compelling projects but also means being ready to challenge certain projects based on the tangible evidence that there is no effect on the beneficiaries.

A good way of raising awareness and explaining impact evaluation is the systematic use of evaluation report presentations to generate open discussions with the various interested parties.

5.1.3 Plan for an annual impact evaluation programme ensuring feasibility and added value

Every year, a certain number of impact evaluations should be planned and appropriate budgets released.

The specific choice of the interventions to be studied should be subject to an explicit case-by-case analysis. Indeed, there is much interest in a quality impact

evaluation but it represents large budgets. The question of value for money must be raised: on the one hand, are the advantages to be derived from the impact evaluation worth the expense? On the other hand, have the minimum conditions for conducting an impact evaluation been met?

The advantages of a rigorous outcome and impact evaluation through the application of mixed methods are as follows:

- Measurement of the effects on the beneficiaries and empirical evidence of their attribution to the project (outcomes).
- Reasoned judgement on the achievement of the outcomes that cannot be measured quantitatively and the achievement of the impact (global objective).
- Understanding of the mechanisms and elements inhibiting or promoting the effects of a project.
- Accountability for various parties (donor, partner, executing agency, beneficiaries) of the effects of an intervention. In the event of a sensitive or disputed intervention or a strategic intervention or simply to account for aid effectiveness, the evaluation's findings can be used by various stakeholders at different levels of detail.
- Decision support. In the case of a pilot project to be extended or extrapolated or repeated projects, the evaluation can guide the decisions based on scientifically demonstrated and/or rigorously reasoned evidence.
- Capitalising on experiences, in terms of the achievement of the objectives of certain projects (summative) and the project or evaluation implementation methodology (formative). With regards the learning process, this is justified when the intervention is particularly innovative and presents a high potential to be replicated.

The costs of an outcome and impact evaluation through the use of a mixed methodology are relatively high. Indeed, in addition to a sometimes significant cost for collecting data, it takes more time (time to collect primary data and time for the quantitative analyses). The costs of an impact evaluation correspond to the methodology to be implemented and therefore the types and quantity of data to be used. Consequently, if the relevant outcome measurements have not been collected and stored on a medium that can be used by the M&E system, the costs of an impact study will be higher. **The cost of an ex post evaluation therefore partly depends on the quality of the formulation of the M&E system and its effective implementation.**

Similarly, the **context in which the evaluation takes place has an effect on data collection costs**: geographic movements, security, the availability of people for interviews, the availability of local people to train teams of researchers and provide a minimum of logistics, the difficulty of collecting certain types of data, etc. are just some of the elements that can have a significant influence on the data collection cost (costs resulting in a relatively long field presence).

Furthermore, these elements that influence the cost, such as the time spent in the field, can also lead to a review of the proposed methodology: focusing on one region (this was the case in **Senegal** and the **DRC**), cancelling primary data collection (as for the projects in **DRC** and **Rwanda**), not collecting certain indicators (abandoning the collection of land productivity data in **Morocco**), restricting the quantitative assessment to one aspect of the project (as was the case for **Senegal** - access to water through new boreholes-).

More fundamentally, the issue of **value for money** affects the place of evaluation and therefore impact evaluation in the current Belgian development cooperation system. As long as the interest and thus the need are not recognised, it is unlikely that the cost/benefit analysis will be positive.

In addition to a cost/benefit analysis to discern the added value of an evaluation, **it is important to ensure that it is feasible**. Thus, there are two other elements to be managed:

- The **nature of the intervention** and the **context of its implementation in the field**. An impact evaluation will have no purpose if the project's logical framework has confused objectives, that are inconsistent and unrealistic with actions and resources implemented in the field, or if the project's implementation has caused many problems, or even if the socio-political environment during the project was unfavourable. In such circumstances, the planned intervention has practically no chance of having any effect. It is therefore pointless spending money implementing a quantitative evaluation system (this proved to be the case for the evaluation of the AETP2 project in the **DRC**).
- **The socio-political environment at the time of the evaluation**. The context in which the impact evaluation is set to take place is another important aspect because it can become impossible to collect the necessary information, or make it very costly, or both. The political environment can hinder an impact study. Indeed, the partner may be reluctant to see the effectiveness of an evaluated intervention. A mixed impact evaluation could constitute a negative judgement about its own policies or because the intervention includes issues that go beyond it and which are of crucial importance for the partner. **The political dimension and challenges of a rigorous evaluation of a project's effects must be considered** (this dimension has likely been underestimated in the case of **Rwanda** for the LRSP project evaluation).

5.1.4 Guarantee the quality of evaluations

Impact evaluation is a complex exercise that combines a research perspective with a development perspective. It is this combination that creates the wealth, value and ultimately its usefulness. It is important to maintain this synergy and take it into account when drafting the Terms of Reference and in the evaluation framework.

The Terms of Reference must **establish the evaluation's objectives; plan an exploratory mission; and anticipate adapting the evaluation to the realities in the field**. The quality of an impact evaluation depends not only on internal factors within the team responsible but also unforeseen but decisive external factors such as, for example, the refusal of the national authorities to support the evaluation.

Constant discussions with the evaluation's originator and a certain degree of flexibility on its part are vital in order to produce a useful, quality work. For example, in the **DRC**, the rather fundamental change in the method (and the team) allowed a rigorous evaluation to be produced, despite the constraints.

Given the use of relatively complex statistical methods, it is important to be able to validate the methodological choices and assess the soundness of the conclusions depending on the data quality, the quality of the counterfactual and the quality of the statistical processes. This can be done internally if there are the skills, if not, this validation must be assigned to an external expert.

5.2 Conduct high-quality rigorous impact evaluations

Conduct high-quality rigorous outcome and impact evaluations by applying a mixed methodology based on the intervention logic.

This study has demonstrated all the benefits of implementing a mixed methodology combining qualitative and quantitative methods with the aim of assessing an intervention's outcomes and impact. The experience confirms the current literature on the subject: it is not a question of choosing one method or another, **but subtlety and appropriately combining quantitative and qualitative methods in the same evaluation process**. Technical mastery (quantitative dimension) along with a good understanding of the intervention logic, the context and the theme addressed by the intervention (qualitative dimension) are essential for producing quality impact evaluations.

By combining qualitative and quantitative methods, the evaluation can quantify a project's effects on the beneficiaries (outcomes) by demonstrating their attribution. This combination supports a reasoned judgement on the achievement of an intervention's global impact. Ultimately, it provides an understanding of the mechanisms and context elements that promote or inhibit these effects.

Remember that the decision to conduct an impact evaluation must be based on a prior analysis of its relevance and feasibility, both in the field and financially (see above 5.1.3). **It is also important to consider its methodological feasibility.**

The methodology of an outcome and impact evaluation is influenced by the evaluation's objectives (accountability, decision support and/or capitalising on experiences), through the time of its implementation (at the same time as the project is implemented, after it has started or when it has finished) and through the available financial resources.

5.2.1 Design the evaluation and project at the same time

It is recommended that the **evaluation methodology for a project's outcomes and impact is designed during the project design stage (ex ante)**. This has a double benefit:

- Firstly, designing an impact evaluation process at the same time as designing the project may have an effect on the quality of the project design itself. Designing the impact analysis prior to a project's implementation requires the evaluation and development practitioners to formulate the hypotheses underpinning the practical achievement of outputs and the achievement of outcomes and impact. This requires a better understanding of the context and a detailed risk analysis. This rigorous ex ante analysis could lead to the review of certain activities or how they are implemented in order to maximise the intervention's chances of success. A better understanding of the challenges, risks and context may also lead to a revision of the intervention's objectives in order to make them more realistic (changing the target values of the indicators for example).

In short, designing an evaluation system at the same time as designing the project allows the logical framework to be used as a real management and evaluation tool.

- Secondly, designing an evaluation system relatively early on in the intervention process allows the implementation of a more rigorous methodology that significantly increases the quality of the quantitative evaluation, which involves:
 - (i) Creating a baseline for relevant and well-defined indicators among a sample of beneficiaries and a counterfactual;

5. Recommendations

- (ii) Establishing a counterfactual by encouraging the sequential implementation of the project (pipeline process) or based on objective criteria considered during the intervention so that it is credible;
- (iii) Collecting quantitative monitoring data on the indicators throughout the intervention and at the end of the intervention, from a group of beneficiaries and ideally a counterfactual and precisely defining the collection arrangements and frequency.

Given the current position of evaluation within the Belgian bilateral cooperation, it is unlikely that the evaluation process for outcomes and impact will be developed and effective in the field from the time a project is implemented. However, the evaluations conducted as part of this study show that it is nonetheless possible to use quantitative methods even when not all optimum conditions are met. **It is therefore recommended that whether the outcome and impact evaluation design is ex ante or ex post, quantitative analysis methods are used as long as it is feasible and relevant to do so.**

Finally, it should be noted that not conducting impact evaluations on the pretext that they reproduce an intervention where evidence of success has been demonstrated is not satisfactory. Indeed, impact evaluations have limited external validity, the aid mechanism and context are determining factors in a project's success, both elements that are difficult to imitate (see above for more detail). **It is therefore recommended that project evaluations are conducted using a mixed methodology respecting the required and minimal conditions.**

5.2.2 Successive steps and important elements in conducting a rigorous ex post impact evaluation

Implementing an evaluation system combining mixed methods involves mobilising a team of experts in qualitative evaluation and the application of quantitative methods.

However, other elements are also important:

- Having a qualified local partner, who understands both the realities in the field and the methodological requirements.
- Finding a field consultant who knows the project and the area, avoiding conflicts of interest. This role must be well-defined from the start so that it does not influence the respondents.
- Providing for the involvement of the various project stakeholders in the evaluation process. It is particularly important to involve them in the preparatory phase of the evaluation, and during the restitution of the evaluation's findings. This helps to address the most relevant questions for those involved, but also allows them to validate the findings. This type of approach can be used to increase the quality and relevance of the study. Furthermore, it is also the opportunity to discuss the methodological process and raise their awareness about the outcome and impact evaluation.

The steps of the evaluation system to be implemented for a quality rigorous ex post evaluation are described below.

- (1) **Preparatory phase.** The objectives are to reconstruct the logical framework, identify the quantitative methodology to be applied and prepare the data collection tools.
- **Documentary analysis to reconstruct the intervention logic:** analysis of relevant existing documents, analysis of the socio-economic, political and cultural context at the time of the project and during the evaluation; analysis of the

theme(s) addressed in the project (through, among other things, academic and scientific literature) and meetings with certain key people in the intervention present in Belgium.

- **Formulation of the hypotheses underpinning the causal relationship leading to the outcomes and impact in the intervention logic and definition of relevant indicators.** The idea here is to highlight the causal relationships between the inputs, outputs, outcomes and impact. This exercise helps to establish any problems in the expected chain of effects. For example, identification of unrealistic outcomes and/or impact in terms of the implemented actions and resources, identification of context elements that are conducive to, or on the contrary, harmful to the proper conduct of the effects. Thus, certain outcomes and/or an impact must be reformulated for greater consistency with the intervention logic. It is at this stage that the relevance and quality of the indicators available in the project's M&E system must be critically reviewed. Lastly, a list of relevant indicators to be collected in the field must be drawn up so as to be able to assess the project's outcomes and impact.
- **Identification of secondary data and creation of information collection tools.** Based on the hypotheses to be tested and the indicators to be collected, it is important to look for relevant data that is already available in the project archives or through national statistics. Data collection tools must then be created, depending on what additional information is required: interview guides (for group surveys or individual interviews with certain key people in the intervention), and/or a questionnaire for a household survey. In an ex post situation, there must be an attempt to reconstruct the pre-intervention situation for the beneficiaries and the counterfactual using data from the respondents' memories or through secondary data.
- **Identification of a counterfactual** to be able to demonstrate the attribution of the project's effects. Based on a good knowledge of the context and the intervention area, a group of non-beneficiaries must be selected that presents similar characteristics to the beneficiary group before the intervention. This step is difficult but not impossible. However, it is important to be aware of two major problems: the contamination effect (spill over effect) and the problem of selection bias (see above in the methodology for a detailed explanation).
- **Definition of samples and the survey area.** Based on the impact survey's objectives (accountability, decision support, capitalisation), the time and financial constraints and secondary data already available, a sample size must be defined that is large enough to ensure any outcomes are detected with a certain degree of precision. The selection of the survey area, as well as that of the respondents must be as random as possible or conducted on the basis of objective criteria uninfluenced by the partner or executing agency (otherwise it could bias the results).
- **Exploratory mission.** Firstly this is the chance to ensure there is approval from the local authorities to conduct an evaluation and their potential support for doing so. Secondly, it is the opportunity to meet certain key people in the intervention living in the country and discover the outputs achieved in the field. Discussions about the relevant issues to be addressed may also take place. It is also the time to collect some secondary data and ensure its reliability. This is when the various information collection tools are tested and the logistics organised for the data collection phase.

The information collected during the exploratory mission is used to revise the intervention logic in light of the realities in the field, review the hypotheses to be tested and refine the indicators to be collected by making corrections to the information collection tools.

- (2) **Field phase.** The objectives are to collect enough good quality data; assemble as many observations and as much information as possible, also through informal discussions. This relatively long time spent in the field is often the opportunity to answer questions from beneficiaries and other project stakeholders, raise problems and have informal debates. It is therefore important to have a well-prepared team of enumerators and take care that everyone sticks to their role.
- Strictly speaking, **primary data collection** only begins after the questionnaire test phase. This step is vital in guaranteeing the quality of the data collected in the field. It consists in ensuring that the enumerators master the tool and that they are competent to perform the required task. The formulation of certain questions can also be reviewed and the knowledge of the realities in the field further refined.
 - The various data collection tools are implemented **under the supervision** of an evaluator who supports the team of enumerators in the field. There are several advantages to the evaluator's presence during the survey phase: dealing with unforeseen circumstances in the field (for example the climate or the state of the roads that prevents certain areas being visited), ensuring that the enumerators do indeed question the people selected, correcting questionnaires as required, providing modifications where appropriate. In addition, through observation and informal meetings the evaluator can build a better understanding of the context and local issues.
 - Finally, the **data must be encoded** into a usable digital format (for example CSPro). Due to labour costs, it is recommended that the questionnaires are encoded locally, but it is essential to have a competent partner that can oversee the monitoring and quality of the encoding (for example a university researcher). It should be noted that the double encoding technique, which consists in two different people encoding the same questionnaire can help to identify encoding errors quickly. However, this is an expensive method. If the survey is conducted on a tablet, there is no encoding phase. The data is automatically available. Nevertheless, there are some disadvantages to using a tablet (recharging in areas without electricity, use by people with little computer training, going back when there are inconsistencies, etc.).
- (3) **Analysis phase.** The objectives are to summarise the results obtained by the quantitative and qualitative analysis, demonstrate the links to the project by explaining the context elements and mechanisms that have promoted (or inhibited in if no effect is observed) these effects.
- **Data cleaning.** This phase is more or less important depending on the quality of the encoding. This work can only be done by someone familiar with the field and the survey phase. It involves ensuring the consistency of the encoded data, not only consistency between the different parts of the same questionnaire but also consistency with the field observations.
 - **Data analysis using quantitative methods.** Firstly, it is important to test the hypotheses underpinning the causal relationships leading to the outcomes by qualifying the observed effects and demonstrating their attribution to the project. To do so, the quantitative data (primary and secondary) is used to conduct statistical and econometric tests (on the STATA software for example). The qualitative data is then used to present the argument for certain causal relationships and assess some of the non-quantifiable effects. Secondly, the results obtained must be explained, i.e. an explanation of the mechanisms that have led to the project's effects or lack of effect. A quantitative analysis offers considerable potential for learning what works and what does not work in development assistance.

- (4) **Restitution phase.**²⁵ The objective is twofold. On the one hand, to raise awareness about impact evaluation starting with a specific case in terms of both findings on projects' effects and evaluation methodology. On the other hand, to validate the findings and/or discuss them with the people concerned.

Once the evaluation reports have been drafted and approved by the originator, it is interesting to organise a restitution in the field with the partner, the executing agency and its beneficiaries. This is a way of realising the added value of a quality impact evaluation compared with any other type of evaluation and generating real debate about aid effectiveness.

5.3 "Management towards outcomes and impact" for more realistic projects targeting the achievement of tangible effects

Incorporate the outcome and impact evaluation into project management in order to design realistic projects that target the achievement of outcomes and impact as a priority. This means incorporating the evaluation at the time a project is formulated and into its implementation as well as developing an effective M&E system to measure the outcomes and assess the impact of a project.

This analysis has shown the weaknesses of the projects, not only in their design, but also in their implementation and their monitoring-evaluation. These weaknesses explain, at least in part, the few positive impacts observed in the three projects examined.

The projects do not prioritise the achievement of outcomes or impact, but focus instead on physical achievements and financial disbursements. In the context of efforts to improve the effectiveness of public spending and strengthening the effectiveness of development assistance, this finding is disconcerting.

Following this ex post impact evaluation of four development projects and on the basis of its conclusions, the report proposes a solution to improve the design/formulation, implementation and monitoring-evaluation of interventions funded by Belgian development cooperation. This aims to improve the achievement of outcomes and impact.

Firstly, it is recommended that the **outcome and impact evaluation is incorporated at the time of a project's design as well as into its implementation and monitoring.** This does not mean that impact evaluations must be conducted for every project. It involves focusing on the achievement of specific objectives and a global objective from the start and not being concerned solely with physical achievements and timely budget disbursement.

There are two advantages to **incorporating the outcome and impact evaluation into project management.**

- The **evaluation process needs to be designed at the same time as a project is designed/formulated.** This significantly improves the quality of the evaluation. More particularly, this helps to design an evaluation system that enables the application of quantitative methods (see above recommendation 5.2.1).
- It may influence the way in which projects are designed and implemented and thus **work towards greater aid effectiveness.** Indeed, incorporating impact at the time of a project's design will mean that unrealistic projects are not

²⁵ In this exercise, a restitution is planned in Belgium, however, no restitution in the field is scheduled.

undertaken and/or are suspended. It will also result in the logical framework being used as a genuine internal management, monitoring and evaluation tool.

In order to make that possible, it is recommended that positive incentives are created at all levels (DGD, BTC, partners, and beneficiaries).

5.3.1 Incorporate impact evaluation at the time of the design and strengthen the project validation process so that unrealistic projects are not started

Incorporating impact evaluation from the time of the design allows a detailed consideration of the factors that can inhibit or promote an intervention's success. This means formulating the hypotheses underpinning the entire causal sequence of a project's intervention logic and consequently providing a better knowledge of the specific context, the challenges at all levels and the risks. Such work beforehand then allows projects to be designed/formulated in line with the realities in the field.

The evaluations performed show that there are several elements affecting a project's design that need to be improved:

- Knowledge of the context, the challenges at different levels (political, social, economic, upstream and downstream of the sector affected by the project) and their repercussions on the proper execution of the project;
- The formulation of global and specific objectives that are realistic, well-defined, quantified (target value to be achieved) and measurable by indicators for which the initial value is known;
- Risk analysis and an objective assessment of potential repercussions on the project;
- A systematic review of the literature (including academic) to ensure that lessons learned elsewhere are taken into account.

Incorporating impact evaluation from the time of the design is a way of improving a project's design, but alongside this, the **process of validating projects before their launch needs to be strengthened**. Although this process exists in the form of a quality control committee, it is not effective enough because it is letting through projects where there is a low probability of achieving the desired objectives. Indeed, a detailed ex ante study would have shown this for the four evaluated projects.

This demands strong commitment, sufficient resources for preparing actions, suitable organisation between the DCGD and the BTC, but above all, the right incentives so that every party works on improving this.

5.3.2 Practice genuine results based management (outputs but also outcomes and impact)

In the projects examined, the M&E systems have been effective for the operational monitoring of the achievement of outputs but they have not generated any relevant usable data for assessing the outcomes and impact of projects. Even faced with the difficulties in executing the activities, **decisions have always favoured disbursement**, and do so, even though it seems obvious that the achievement of these outputs could not lead to the expected outcomes.

Incorporating outcome and impact evaluation at the time the project is designed and implemented can help to establish a genuine results based management system. It is therefore important to design an effective and operational M&E system to monitor the achievement of outputs, but also monitor and evaluate the achievement of outcomes (and impact). It is not enough to record the achievements, we must also remain critical about the outputs achieved (are they used? by whom? for what?).

Management based on outcomes and impact (management towards results)

requires a M&E system that can

- Report regularly on the project's progress (status of achievements and disbursements) and causes for delays or potential problems;
- Account for shortcomings in the achievement of outcomes and identify the causes for this;
- Provide information about worrying aspects of the project or concerns about achieving the objectives;
- Provide a decision-making ability to challenge or review a project which is no longer in line with the desired outcomes and expected impact; and therefore potentially decide to suspend or cancel an intervention where necessary.

For an **effective M&E system to be operational**, the **methods for collecting information** must be clearly defined: who collects what data, how frequently, on which medium, with what support, with what objective? Furthermore, it is vital to **create the right incentives at every level** (DGD, BTC, partner, beneficiaries) without which, it cannot be operational due to a lack of means and/or time and/or human resources.

6. Bibliography

Bamberger, M., 2012. Introduction to Mixed Methods in Impact Evaluation. Impact Evaluation Notes, Rockefeller Foundation & Interaction.

DAC, 2010. Glossary of Key Terms in Evaluation and Results Based Management.

Deaton A.S., 2009. Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development. Working paper 14690. National Bureau of Economics Research.

Gertler P.J. et al., 2011. Impact Evaluation in Practice. World Bank.

IEG (Impact Evaluation Group), 2013. Impact Evaluation-the experience of the independent evaluation group of the World Bank.

IEG (Impact Evaluation Group), 2013. Conducting quality impact evaluations under budget, time and data constraints. World Bank.

Naudet J.D., Delarue J., Bernard T., 2012. Impact Evaluation: a Tool for Accountability? Lessons from experience at AFD. AFD working paper.

White, H., 2009. Theory-based Impact Evaluation: Principles and Practice. Working paper 3, 3ie.

White, H., 2013. An introduction to the use of randomized control trial to evaluate development interventions. Working paper 9, 3ie.

Appendices

APPENDIX 1: EX POST IMPACT EVALUATION REPORT ON THE AETP2 PROJECT IN THE DRC

APPENDIX 2: EX POST IMPACT EVALUATION REPORT ON PARPEBA IN SENEGAL

APPENDIX 3: EX POST IMPACT EVALUATION REPORT ON THE PMH PROJECT IN MOROCCO

These reports are not printed and are only available in French on the CD-ROM in appendix of the synthesis of the report or on the website.



KINGDOM OF BELGIUM

Federal Public Service
**Foreign Affairs,
Foreign Trade and
Development Cooperation**

Office of the Special Evaluator for International Cooperation

Egmont • Rue des Petits Carmes 15, B-1000 Brussels • + 32 2 (0)2 501 38 34 • www.diplomatie.belgium.be • www.dg-d.be • ses-dbe@diplobel.fed.be

Legal registration : 0218/2014/003