RUHR
ECONOMIC PAPERS

Gunther Bensch
Jörg Peters
Christoph M. Schmidt

Impact Evaluation of Productive Use – An Implementation Guideline for Electrification Projects

Productive Use of Electricity (PRODUSE)
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)
Africa Electrification Initiative (AEI)

RUHR UNIVERSITÄT BOCHUM RUB RWI #279

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Joachim Schmidt
RWI, Phone: +49 (0) 201/81 49-292, e-mail: joachim.schmidt@rwi-essen.de

Ruhr Economic Papers #279

Responsible Editor: Christoph M. Schmidt

ISSN 1864-4872 (online) – ISBN 978-3-86788-324-5

Ruhr Economic Papers #279

Gunther Bensch, Jörg Peters, and Christoph M. Schmidt

Impact Evaluation of Productive Use – An Implementation Guideline for Electrification Projects

Productive Use of Electricity (PRODUSE)
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)
Africa Electrification Initiative (AEI)

RUHR UNIVERSITÄT BOCHUM RUB
RWI

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

ISSN 1864-4872 (online)
ISBN 978-3-86788-324-5

Gunther Bensch, Jörg Peters, and Christoph M. Schmidt[1]

Impact Evaluation of Productive Use – An Implementation Guideline for Electrification Projects

Abstract

There is a consensus in the international community that rural electrification and, in particular, the productive use of electricity contributes to poverty alleviation. At the same time, efforts to evaluate the impacts of development projects have increased substantially. This paper provides a hands-on guide for designing evaluation studies regarding the impacts of productive electricity usage. Complementary to the existing literature on evaluation methods, this guide familiarizes project managers with the concrete steps that have to be undertaken to plan and implement an evaluation. The guide comprises three modules based on enterprise surveys and on anecdotal case studies. For each module, the implementation is described on a step-by-step basis including conceptual issues as well as logistics and methodological questions.

JEL Classification: O22

Keywords: Development effectiveness; productive electricity use; survey design; monitoring and evaluation

August 2011

1 *Gunther Bensch, RWI; Jörg Peters, RWI; Christoph M. Schmidt, RWI, RUB, IZA Bonn, CEPR London.* *– Financial support by World Bank's Africa Electrification Initiative (AEI) is gratefully acknowledged. The development of this guideline is part of the joint research project Productive Use of Electricity (PRODUSE) of Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and World Bank. We are grateful for valuable comments and suggestions by Anna Brüderle, Kilian Reiche, and Colin Vance.* *– All correspondence to Jörg Peters, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Hohenzollernstr. 1-3, 45128 Essen, Germany. E-Mail: joerg.peters@rwi-essen.de.*

1. Introduction

Rural electrification is widely considered to be a prerequisite for development and the removal of barriers hampering economic growth: Beyond the usage for household activities such as lighting or television, it is expected that electricity increases the productivity of enterprises. There is a consensus among practitioners in electrification projects as well as energy experts in international cooperation partner countries (ICPC) and donor organisations that the productive use of electricity (PU) is key for the sustainability of poverty-alleviating impacts of electrification projects.[1]

At the same time, the international community has increased efforts to evaluate project impacts in order to improve the accountability of development projects. International initiatives such as the International Initiative for Impact Evaluation (3ie) or the Network of Networks on Impact Evaluation (NONIE) promote rigorous evaluations. RAVALLION (2008) and GERTLER ET AL. (2010), for example, elaborate methodological recommendations for these rigorous evaluations. With a particular focus on electrification projects, PETERS (2009) proposes different approaches to identify project impacts.

While these publications appropriately summarize the methodological part of an evaluation project, they are either targeted at academic evaluation researchers or at practitioners with a high affinity to becoming acquainted with evaluation methods. Practitioners who are rather interested in setting up a hands-on monitoring and evaluation (M&E) scheme to obtain robust insights into the impacts of small-scale intervention can hardly be expected to familiarize themselves with these methodological issues at the level of highbrow econometric research, though.

Intending to close this gap, this paper provides guidance on how to design an Impact Evaluation System. The particular focus is on PU in electrification interventions, examining electricity take-up and income generation in small and micro enterprises (SME). We address this guideline to those people in development agencies, regional energy utilities, or local

[1] Our definition of "productive use of energy" follows the working definition by the Global Environment Facility (GEF) and UN Food and Agricultural Organization (FAO): "In the context of providing modern energy services in rural areas, a productive use of energy is one that involves the application of energy derived mainly from renewable resources to create goods and/or services either directly or indirectly for the production of income or value" (White 2002a). Impacts are defined according to OECD/DAC (2002) as "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."

research units in charge of commissioning a PU impact evaluation (whom we – for simplicity – hereafter call "project managers"). Still, it is in the same way geared towards researchers or practitioners in charge of the evaluation itself, the "researchers".

The education of this audience with respect to methodological issues is not the focus of this discussion. Rather, its major aim is raising awareness for important parameters in the design of a PU Impact Evaluation System and the provision of project managers with an accessible menu of requisite steps also intended to encourage the further development of local evaluation capacities. While the application in the focus of this article is evaluating the impacts of electrification on SMEs, the principal steps of the proposed Impact Evaluation System are interchangeable and can be transferred to other development projects.

In order to stress the demarcation between a classical Monitoring System and an Impact Evaluation System, the guideline reviews briefly the different results of an intervention: outcomes, intermediate impacts, and highly aggregated impacts. Classical Monitoring Systems typically address project activities and sometimes outcomes, but not impacts (ADB 2006). This is elaborated in Section 2 – also by discussing the problems and pitfalls that one encounters when the impacts of electrification on SMEs are to be evaluated.

Section 3 first introduces principal strategies to assess the impact of an intervention. Subsequently, three modules are presented: One simpler quantitative module based on a short SME survey (Module A), one extended quantitative module based on a profound SME survey (Module B), and one qualitative module based on anecdotal case studies (Module C). All modules have been field-tested intensively. For instance, they were applied in the course of the research project *Productive Use of Electricity* (PRODUSE), commissioned by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and the World Bank.[2] A discussion of their respective opportunities and limitations complements the proposal of the modules.

Section 4 is the core of the guideline and presents the process of designing a PU Impact Evaluation System in a step-by-step way. Step 1, "Getting Started", pays particular attention

[2] The Impact Evaluation approaches presented here were applied as part of the PRODUSE study in Benin, Ghana, and Uganda. The authors implemented comparable evaluation studies in Burkina Faso, Benin, Indonesia, Rwanda, Senegal, and Mozambique. In addition to the PRODUSE report, published reports are BENSCH AND PETERS (2010), BENSCH, PETERS, AND SCHRAML (2010), and HARSDORFF AND PETERS (2010). Methodologically more elaborated methods are used, for example, in BENSCH, KLUVE AND PETERS (2011) or PETERS, VANCE, AND HARSDORFF (2011).

to outlining the decision process, Step 2 describes the designing process of the study, Step 3 the survey preparation, Step 4 the implementation of the survey, and Step 5 the data analysis and the reporting. To facilitate its practical applicability, the guideline provides links to sample questionnaires that have been used in the PRODUSE study and other evaluations. The concluding Section 5 distils the lessons to be learned from the previous sections, making the illustrative example of the PRODUSE study a valuable blueprint for other projects.

2. Classical Monitoring vs. Impact Evaluation

2.1. Outcomes, impacts and highly aggregated impacts

Any programme implemented in practice aims at making a genuine difference to the state of well-being in the target population. To this end, the programme directly influences the state of *outcome* variables that are intended to trigger *intermediate impacts* and, eventually, *highly aggregated impacts* on income, nutrition, or other variables of fundamental importance. In the case of productive electricity usage, one might consider the example of a programme that subsidizes the extension or densification of the national grid. Here, a results chain, which connects the intervention's inputs and activities to its outcomes and impacts in generic terms, would consist of the following links: The desired outcome with regards to productive use is that SMEs get connected and use electricity for their production process. An intermediate impact is the effect that this electricity usage has on the firm's production process (increased productivity), while the highly aggregated impact occurs at the level of the firm owner or the firm's employees in the form of higher incomes.

Outcomes are typically clearly attributable to the project's intervention. Both intermediate and highly aggregated impacts, in contrast, might be caused by different impulses. Apart from the project's intervention, such other factors may be the firm's development along its secular growth path, general economic growth, or changes in market prices of the firm's products. In the project's results chain, this insight is expressed as the so-called *attribution gap* between outcomes and impacts. Before attempting any quantification of either of them, the careful enumeration of what the outcomes and impacts are and what the project could achieve in principle should be the starting point of every evaluation effort.

The results chain also shows the difficulty of an Impact Evaluation System: On the one hand, only the highly aggregated impact variables are of ultimate interest when gauging the effectiveness and success of the programme. The intermediate impact variables, *higher profitability*, for example, are no means to the end. On the other hand, the more aggregated the impact indicator is, the more difficult and costly it is to to isolate the *net effects* of the intervention on the impact indicators. Taking our example, these are the effects of the electrification project alone. *Gross effects*, in contrast, also include influences due to external factors that would also have taken place in the absence of the project. Disentangling the electrification impact from such other influences is much more difficult for highly aggregated impacts than for intermediate impacts. In other words, the attribution of causes and effects becomes more difficult for them.

Therefore, the question of which level of results to monitor and evaluate is a crucial question to be addressed by the electrification project's managers. In this spirit, an *impact evaluation* intends to go beyond the demands of a classical monitoring system by also investigating the indirect benefits (impacts) of the intervention. A classical monitoring system, by contrast, is basically restricted to tracking progress of programme implementation and to the review of achievements of the programme's intended direct benefits (outcomes). If it is decided to look into both outcomes and impacts, the present guide provides a pragmatic outline on how to design the implementation of an PU Impact Evaluation System.

The approaches described in this guideline mainly aim at intermediate impacts such as higher profitability or firm creation. Since, for instance, entrepreneurial activity is a promising avenue to economic development, these intermediate impacts can be considered as a prerequisite and, thus, as proxies for highly aggregated impacts. While there is certainly no guarantee that intermediate impacts will ultimately translate into highly aggregated impacts, convincing evidence for the presence of intermediate impacts is an important piece of information when assessing the possibility that the programme has induced positive highly aggregated impacts or not. Intermediary impacts can, hence, be seen as "stepping stones" in the endeavour to identify the genuine impact of the intervention on the ultimately meaningful dimensions of people's well-being.

2.2. Second round effects

Even if the net effect of electrification on connected firms (the micro-effect) can be isolated successfully, this is only one step towards a meaningful assessment of the programme's impact. In order to obtain the beneficial effect on the local economy as a whole (the macro-effect), one needs to account for so-called second round effects. The most important second round effect is the crowding out effect. Crowding out effects occur if the benefit of one enterprise is at the expense of other enterprises. For example, if a small shop attracts more customers thanks to its new electric light bulbs, it is straightforward that other non-connected shops lose correspondingly, because their old customers now buy at the connected shop.

In principle, the intervention area as a whole only benefits, i.e. the macro-effect is only positive, if (i) productive electricity usage replaces imported goods by locally produced ones, or (ii) new goods for export are produced using electricity, or (iii) the total productivity of the local economy increases. For instance, an overall productivity increase might be a result of intensified usage of mills instead of mortar and pestle, liberating productive capacities for other purposes.

While it is difficult to fully account for such crowding out effects, they have to be kept in mind in both designing a PU Impact Evaluation System and interpreting its findings. At minimum, an attempt should be made to obtain indicative evidence for such effects. This could be achieved, for example, by including non-connected SMEs in the PU Impact Evaluation System as control group or by qualitatively probing into the question of where the customers of newly connected enterprises are attracted from.

Further second round effects are possible. Budget effects, for example, occur if people in a village spend parts of their limited budget on new products (e.g. photocopies, cold drinks) that were not available before electrification. As a consequence, they reduce their expenditures on products they used to buy before electrification. This becomes very evident in the case of expenditures for electricity itself – in many cases a typical "imported" good. People no longer buy their candles at the local shop, thereby shifting parts of the value added out of the region.

3. Compiling a Productive Use Impact Evaluation System

The objective of a PU Impact Evaluation System is obtaining credible evidence for impacts of the project – taking into account the attribution difficulties described above. This means that considerable effort has to be designated to disentangle net project effects from gross effects. To successfully address different levels of impacts, the guideline presents different modules for PU Impact Evaluation Systems, together with the required resources.

3.1. General strategies to isolate the project's net effect

The methodological challenge of any Impact Evaluation System is to isolate the net effects of an intervention and thereby attribute changes in indicators to their underlying causes, specifically to the intervention whenever this is appropriate. For this reason, the analyst has to identify the *counterfactual situation,* which is what would have happened to the beneficiaries' (e.g. connected SMEs') relevant outcome variables (e.g. revenue) in the absence of the electrification intervention. Comparing the counterfactual situation to the factual situation – what has actually transpired after electrification – provides a valuable assessment of the true impact of the project. As a matter of course, however, the counterfactual situation is unobservable: we can never know for sure what change would have occurred among the beneficiary group if the programme had not been implemented, and the programme impact can at best be estimated in a convincing fashion.

To find such a convincing estimate, we have to plausibly approximate this unobservable counterfactual situation. In practice, three main so-called identification strategies exist: i. Simple before-after comparisons (the same firms are interviewed before and after electrification), ii. simple cross-sectional comparisons (connected and non-connected firms are interviewed at one point in time) and iii. before-after comparisons with control group (all firms are interviewed before electrification; some of the firms connect, and these connected and non-connected firms are interviewed again after electrification). These three principal strategies differ in their ability to fend off various problems arising from unobservable heterogeneity and, correspondingly, in their informational requirements and associated cost.

An in-depth explanation of these identification strategies with a particular focus on electrification projects can be found in PETERS (2009). This article includes a discussion of the

assumptions under which each strategy is able to obtain a reliable estimate of the net effect. In a nutshell, a before-after comparison will only be able to isolate the genuine impact of an intervention, if there is no underlying trend in the relevant outcome variables that would occur also without the programme. Similarly, a simple cross-sectional comparison is only a reliable way to isolate the programme impact, if connected and non-connected firms do not differ systematically in terms of unobserved characteristics, which tend to influence the relevant outcome variables. And the combination of these two approaches requires that the underlying trends for both, the connected and the non-connected firms, are identical.

3.2. Three PU impact evaluation modules

In any programme's results chain, the range from outcomes to highly aggregated impacts typically offers an escalation of overlapping influences and, correspondingly, of informational demands for isolating net effects. Since the evaluation problem might become prohibitively difficult to solve as the questions asked increase in terms of their ambition, practitioners need a menu of approaches to move along the results chain as far as the particular application allows.

Following this reasoning, we propose three modules for PU Impact Evaluation that are tailored to measure impacts in the context of productive electricity use and that have been field-tested in various developing countries: MODULE A, based on a short enterprise survey, MODULE B, based on a profound enterprise survey, and MODULE C, a case study approach based on targeted qualitative enterprise interviews.

In our assessment, evaluators should pursue one of the two quantitative research approaches, MODULE A or B, where the concrete choice between them depends on both cost and ambition. Moreover, MODULE C serves as a qualitative complement, which enables researchers to draw conclusions beyond the realm of quantitative analysis. As WHITE (2002b) points out, this combination of qualitative and quantitative approaches can serve as one remedy to their respective limitations. In a nutshell, the advantage of qualitative research is the open way in which interviews are conducted compared to the corset of a structured questionnaire in quantitative surveys. Yet, qualitative evidence can hardly be more than anecdotal without huge financial efforts. This also leads to the primary advantage of

quantitative research, which enables the researcher to average across many observations, thereby benefiting from the law of large numbers.

In contrast to the profound survey, the short enterprise survey aims at "easy to get and handle" information and abstains completely from eliciting highly aggregated impacts such as profits or improvements in market access. The aim of this modesty is to avoid any misleading findings on more complex issues that might result if no sufficient methodological effort is dedicated (e.g. with regard to sample size or advanced statistical data analysis). After all, estimating highly aggregated impacts requires the convincing construction of counterfactual situations, a highly information-hungry affair. Following a more modest approach, MODULE A envisages providing evidence on outcomes and on impacts that are close to the attribution gap. Consequently, one needs to appeal to plausibility when linking the observed changes in the direct results and impacts of the intervention to highly aggregated impacts. If the survey, for example, shows a considerable take-up of machinery, one might plausibly assume that this also positively affects productivity and, hence, firm profits and employees' wages.

MODULE B, the profound survey, by contrast, aims at providing direct evidence for such effects. Since the credible construction of the counterfactual, no-intervention situation, is at the heart of any impact evaluation, evaluators need a plethora of information. For instance, if a plausible counterfactual situation is established, the impact of electrification on firm profits can be assessed by comparing the electricity-using firms to their counterfactuals.

MODULE C, the targeted qualitative enterprise inquiry, is included, since SMEs are less homogenous and numerous than households, making a quantitative analysis more difficult. For example, only one or two larger firms might exist in a target region. Including them in a quantitative study would not be reasonable, since the advantage of quantitative studies – taking the average across many observations – could not be exploited. Conducting case study-like interviews is much more sensible in this case. Another reason for applying the targeted qualitative inquiry is accounting for unintended effects or probing more deeply into certain issues than would be possible with structured questionnaires, for instance crowding out effects as delineated in Section 2.2. Thereby, such case studies can help understanding difficult processes emerging among beneficiaries, and provide for anecdotal evidence that can, not least, be fed into the design of future quantitative studies.

Table 1 catalogues the main features of the three modules – including their respective opportunities and limitations. Of course, the components can be modified for specific reasons and the different parts of the three modules can be combined. Based on our experience in various projects, we believe that the modules are a reasonable compilation of features in order to yield the described results and to help the project managers designing their PU Impact Evaluation System along these lines.

Please note that although the MODULE B would be commonly referred to as the "rigorous" way of doing M&E, this term is purposefully avoided. The reason is that, as WHITE (2002b) points out „… the real basis for *rigor* is the proper application of techniques. Badly or misleadingly applied, both quantitative and qualitative techniques give bad or misleading conclusions." In this sense, all modules proposed here can and should be applied rigorously.

Table 1 : Potential approaches for PU Impact Evaluation – An overview

Module A Short Enterprise Survey	**Module B** Profound Enterprise Survey	**Module C** Targeted Qualitative Enterprise Inquiry
Main purpose		
Providing evidence on impacts close to the project's direct outcomes that do not need an extensive survey set up for causal attribution. Relation to ultimate poverty impacts is instead established on a plausibility basis only, by a discussion of results chains.	Providing evidence on the causal relationship between electrification and ultimate development indicators, using state-of-the-art evaluation techniques.	Collecting anecdotal evidence on electricity usage and its impacts. Its focus lies on issues that can hardly be addressed in structured interviews or on very rare SME types (e.g. large firms).
Identification Strategy	*See PETERS (2009) for more details.*	
Before-after comparison	Cross-sectional, before-after comparison or before-after comparison with control group[3]	Before-after comparison or retrospective questions (with critical qualitative assessment)
Sampling Method	*See as well Step 3g, Section 4.*	
Simple random sampling.	Simple random sampling or stratified random sampling.	Simple random sampling or purposive sampling of SME of specific interest. If combined with another module, firms can be selected according to stylized firm types detected during the surveys.
Sample Size	*See as well Step 3f, Section 4.*	
Small sample (50-100 SMEs)	Larger sample (>300 SMEs)	5-20 selected SMEs

[3] The baseline survey in a before-after strategy additionally allows gathering profound knowledge about the target region. It can be particularly interesting from the project's perspective to include an already electrified control region. This allows the project to gain insights about what can be expected concerning the behaviour of the rural population/enterprises after electrification (see PETERS 2009 for methodological details of this approach).

Covered Indicators		
Direct outcomes of the intervention. Additional indicators on project-relevant questions can be added. Collected information has to be - easy to determine by respondent - relatively insensitive to formulation of questions - unaffected by an auspices bias[4] - easy to quantify and process.	All indicators of the short enterprise survey are integrated in this module. In addition, the more detailed questionnaire allows for gathering the more-difficult-to-obtain information. For example, detailed questions on sales, raw materials, labour and capital input avoid sensitivity and auspices biases in assessing firm income.	Open-ended questions provide the opportunity to follow unexpected threads in the interview, e.g on reasons for connecting or not connecting, or market access barriers. Indirect and second round effects are also brought up, e.g. if the respondent is aware of competitors who have not benefited from the intervention.
Questionnaire	*Refer to Electronic Annex 1 - 3 in the online version of this article for sample questionnaires or interview guides.*	
Structured, but short; focused on easy-to-get-information. Interview length around 30 minutes.	Structured, covering all dimensions of firm activity, accounting for seasonality, decisive variables such as employment or firm profits are addressed in more detail and multiple ways in order to allow cross-checking. Interview length ca. 60 minutes.	Open; interview guideline should be pursued while leaving space for spontaneous, discursive deviations in directions indicated by the respondent. Interview length 30-120 minutes.
Information Processing		
Simple data analysis with Excel data entry sheet to be prepared by researcher.	Statistically advanced data analysis using statistical software (SPSS, STATA etc.)	Systematic analysis of interview notes along the lines of the underlying guiding questions.
Implementation		
Can be implemented by own project staff, interns, or consultants without particular skills in evaluation methods or statistics; supervision/ backstopping by evaluation researchers is recommendable.	Profound skills and experience required in all stages, i.e. survey design and implementation as well as data analysis; some background in development (and electrification) projects and the respective country recommendable; for data collection, backstopping of experienced local enumerators by methodologically skilled researchers.	Should be implemented by or under close supervision of lead researcher; recommendable to hire consultants familiar with (qualitative) evaluations.

4. Step-by-step towards an effective PU Impact Evaluation System

The project manager might scrutinize the demands of the project, choose an appropriate identification approach (see 3.1) and apply it using one of the three modules (see 3.2.). But what is the best sequence of making these choices and which are the questions to be

[4] Auspices bias (also called courtesy bias) refers to the frequently observed tendency of an interviewee to respond something the enumerator (does not) like(s) to hear. For example, an entrepreneur in a connected firm might answer more positively in an electrification project's impact survey, because s/he is thankful for the electrification. Likewise, in certain question blocks, s/he might give negatively biased answers because s/he expects more funds from the project.

addressed systematically in this process? This section discusses the practical issue of designing a PU Impact Evaluation System, suggesting, which of the different stakeholders of the project should be integrated in which of the stages of the process. Steps 1 and 2 have to be conducted by the project managers or at least require their close involvement. Steps 3 to 5 are mostly the responsibility of the project staff members or of the external researchers to whom the implementation of the PU Impact Evaluation System is assigned. In order to complement the guidance and information provided here and for further readings, one may consult, for example, NONIE (2009) and GERTLER ET AL. (2010).

Step 1: Getting started

Before thinking about the concrete design of the PU Impact Evaluation System in step 2, the project manager should take the following basic considerations.

Step 1a: Decision on whether to do an impact evaluation

Are the additional benefits of a PU Impact Evaluation System compared to classical Monitoring System useful for the project? Do these additional benefits justify the additional costs from the project's perspective? If yes, continue with Step 1b.

The intention of doing a particular PU Impact Evaluation System should be communicated to all other project stakeholders including local partner institutions on both the political and implementation level (e.g. utilities, ministry). They should be included, in particular, in the designing process, if available.

Step 1b: Examination of the project's results chain

The project's results chain is the conceptual framework of the PU Impact Evaluation System. If no results chain with regards to productive electricity use has been established, it has to be drafted in order to get a clear picture of which transmission channels from inputs to impacts are expected. Accordingly, the results chain helps to determine appropriate outcome and impact indicators. Even if a results chain has already been established, a review is recommendable at the time the PU Impact Evaluation System is designed, not least since adaptations in the project design might have occurred in the meantime.

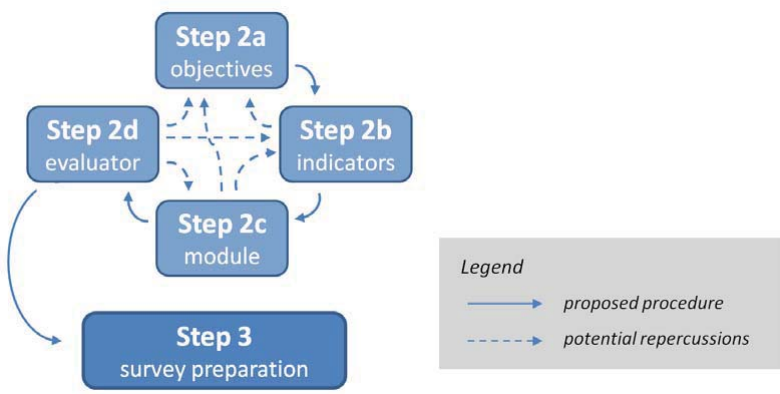
Step 2: Designing

The second step is then to design the PU Impact Evaluation concept. This includes the following parameters:

- Determination of the objectives of the evaluation (Step 2a)
- Decision on the impact indicators (Step 2b)
- Choice of the appropriate PU Impact Evaluation module (Step 2c)
- Selection of researchers to implement the PU Impact Evaluation System (Step 2d).

As depicted in Figure 1, decisions on a certain sub-step may have repercussions on previous sub-steps. For example, if it is decided on Step 2d to hire an external researcher, a revision of previous steps, indicators to be examined is reasonable. Likewise, the decision on which module to apply (Step 2c) can also affect the selected indicators (Step 2b).

Figure 1: Steps in the Design of the PU Impact Evaluation System



Source: Own illustration.

Step 2a: Determination of the objectives and scope of the PU Impact Evaluation System

The first step in designing a project-specific PU Impact Evaluation System is to agree on its objectives. The crucial point here concerns the ambition, i.e. how far the evaluation shall cover the results chain. Does the project want to monitor connected firms and the usage of electricity only or also address higher-level impacts like profits or employment?

The chosen objectives may be subject to changes when deciding on the characteristics of the PU Impact Evaluation System later on, as indicated in Figure 1 above. For example, this can be the case, if budgetary restrictions turn out to impede the implementation of a more sophisticated method (Step 2c) or if indicators considered as indispensable in Step 2b make it necessary to reconsider the objectives of the PU Impact Evaluation System.

Step 2b: Decision on impact indicators

Indicators are direct and unambiguous measures related to the intended goals of a project. Indicators for the evaluation of impacts on productive electricity use range from simply counting the number of connected firms and the appliances they use to their profits, the number of employed workers and the wages they earn. Based on these indicators, concrete questions are to be formulated for the questionnaire. The choice of indicators has clear implications for the module to be chosen in Step 2c (see also Section 3.2). For example, the indicator "Used appliances" can be checked with less effort (i.e. MODULE A) than "Firm profits" (for which MODULE B is required).

GTZ (2007: p.14ff) delineates aspects to be taken into account when constructing project-specific indicators. Such guidelines are important to follow in order to attain a priori neutral indicators that reliably record the degree of progress in the achievement of the proposed results. M&EED Group (2006) lists a range of potential indicators applicable to productive use of electricity. Potential impacts that have not been intended by the project – be they positive or negative – should also be considered and included by appropriate indicators. For all chosen indicators, it should be checked at this stage whether relevant data could be obtained from other sources. This includes official statistics, but also baseline data from other projects or the project itself.

Most indicators require an interview with the firm owner. Some impact indicators may necessitate further interviewees, for example, in order to obtain the perception of employees on the impact of electricity on their working environment. Such research questions, though, are best included in complementary qualitative interviews conducted in MODULE C. Another example could be the impact on the community in total, on the local environment, or the choice of the electricity source: Mini-grids fed by diesel generators, for

example, may result in high long-term costs and dependency to external suppliers, whereas micro-hydro projects may interfere in the local water provision of households and farmers.

Step 2c: Choice of the evaluation module

One of the three modules proposed in Section 3.2. has to be selected: The Short Enterprise Survey, the Profound Enterprise Survey and the Targeted Qualitative Enterprise Inquiry. The module decision should be based on a comparison of opportunities and limitations of each module (refer to Table 1) with the objectives of the evaluation (see Step 2a) and the available budget. Modifications of the selected module can be carried out in line with particular needs of the project. An extensive calibration should be done by the staff member or consultant to whom the implementation of the PU Impact Evaluation System is assigned to in the following Step 2d.

Step 2d: Assignment of implementation to qualified staff members or external experts

The different modules require different levels of skills and resources. The module presentation in section 3.2 indicates the requirements in terms of methodological know-how and time requirements to implement each module. If it is intended to apply econometric methods during the data analysis for MODULE B, the researcher should be familiar with statistics and econometrics – at best documented by a list of academic publications in the fields of impact evaluation and applied econometrics. In case external consultants or researchers are contracted, previous steps in this procedure can be revised by them.

Step 3: Survey preparation

Survey preparation varies substantially between the different modules. For MODULE A, the sub-steps of this task do mostly not apply, since its features are already pre-defined, e.g. the before-after approach is the only recommended identification strategy (step 3a) and no control regions are to be included (step 3e). MODULE B and C, in contrast, require considerably more effort both with regard to desk work and studies (steps 3a to 3d) and to field work (step 3e to 3h). The field work implies a mission of the researchers to visit the target and potential control regions, but also to meet the project staff (in particular if the

researchers are international experts), to finalize the methodology and to train the survey team.

Step 3a: Decision on identification strategy

As described in Section 3.1 and comprehensively outlined in PETERS (2009), different ways of identifying the impacts of electrification exist. An appropriate comparison to the electrified SMEs, the so-called counterfactual situation, has to be established. If the PU Impact Evaluation System is set up at the beginning of the electrification project, in principle all strategies are possible. If the decision to evaluate impacts is taken after the project has electrified the target regions, only a cross-sectional approach is possible. As a matter of course, the before-after comparison requires more resources, since two surveys have to be conducted (before and after), potentially even in two regions (project's target region plus control region, see Table 1).

Step 3b: Submitting an inception report

An inception report should be drafted by the researchers to briefly outline the intended procedure at the outset of the assignment. It provides an opportunity for the project staff to get acquainted with the intended approach and to intervene if deemed necessary. The submission of an inception report is, hence, particularly recommended in case the researcher is an external person or entity, but can also be a valuable preparatory instrument for in-house discussions.

This inception report should best be structured as follows: (i) project basics and conditions, (ii) methodology and (iii) implementation. The first section should present basic information on the electrification project including its results chain. The second chapter should first briefly explain the selection of modules. In a second step, adaptations to the chosen module(s) can be illustrated. The purpose of the third chapter is to present an outline of the data collection and analysis process supplemented by a rough time schedule. This should also include the envisaged sample size – if possible, already specifying the different SME types to be interviewed.

Step 3c: Development of questionnaire

Based on the proposed approach outlined in step 3b, a questionnaire has to be developed that covers the requirements determined in 2a and 2b and that fits into the module determined in 2c. Of course, the questionnaire for MODULE A has to be much shorter than the one for MODULE B. In all cases, the questionnaire should be well organized and furnished with complementary annotations for the enumerator, e.g. if multiple answers are allowed to the specific question.

For MODULE C, the questionnaire is more an interview guideline delineating the purposes that the interview should pursue – in spite of its principal open-ended nature. Sample versions for the three modules are provided in the Electronic Annex 1 to 3 in the online version of this article, where the interview guide for MODULE C refers to the supplementary qualitative run for the PRODUSE study in Uganda, for which findings are presented in NEELSEN AND PETERS (2011).

At least for MODULES A and B, pre-testing the questionnaire with 5-20 interviews is imperative to scrutinize the formulation of questions. The interviews for MODULE C are more conversational and adaptations in the course of data collection are – in principle – possible. It is most suitable to do the pre-testing with the already selected and trained enumerators (Step 3h). At the same time, the pre-test can serve as a training component for the enumerators. It is also highly recommendable for the researchers to check for the appropriateness and completeness of the questionnaire at the beginning of the survey preparation mission by doing field trips to the target region and through a few focus group discussions with target group representatives.

Step 3d: Selection of field work team

For MODULE A the team may even consist of project staff only. Additionally, interns or consultants can be hired. By contrast, MODULE B requires one or two teams of around four enumerators and one field supervisor, depending on the sample size and availability of time and means of transport, of course. As a rule of thumb, one can expect 4 and 6 interviews per enumerator per day for MODULE B and A, respectively. Interviews for MODULE C should be conducted by the hired researchers themselves, supported by local consultants familiar with the situation and social mores in the target region.

Step 3e: Choice of control regions

Information that allows assessing the comparability of potential control regions and the target region of the electrification project should already be collected as part of the preparatory desk work. In addition, a field trip to the target areas of the intervention is generally indispensable. While the comparability of villages can best be assessed on the ground by visual inspection, the following list of criteria can provide for some guidance:

- distance to the capital and/or regional centers;
- population size;
- main source of income (agricultural and non-agricultural products);
- road accessibility (distance to asphalt roads; accessibility by cars and/or trucks);
- transit traffic;
- existence of a regular market in the village;
- political relevance;
- presence of other development projects.

Talking to local key informants like village chiefs, teachers, or NGO representatives can help to get a picture of the villages potentially to be included.

Step 3f: Determination of sample size

The determination of the sample size for MODULE A or B, in principle, is based on statistical considerations. However, a statistically accurate determination of the required sample size, commonly referred to as *power analysis*, will not be possible in most cases. This statistically appropriate sample size mainly depends on the specific impact indicators (e.g. firm profits or employment, lighting usage) and the extent to which they are expected to change due to electrification: The smaller the expected change, the higher the sample size that is required to derive robust and clear interpretations from statistical results. In a nutshell, if one finds statistically significant evidence for an impact of electrification on, for example, firm profits, there are not so many reasons to worry about a sufficiently large sample size. The problem is rather whether to interpret a no-effect result as genuine evidence of no effect of the intervention or as a reflection of an insufficient sample size, given the setup of statistical significance tests. It might as well be the case that the sample size is simply too small to

detect a positive impact. The objective of a power analysis is exactly to avoid such inconclusiveness.

See, for example, MAGNANI (1997) for an accessible presentation of power analysis.[5] Among the parameters required to determine the sample size, are (with the + or - indicating whether the parameter increases or decreases the required sample size):

a) the number of firms in the target population [+];
b) the heterogeneity of firms in the target region [+];
c) the expected magnitude of the intervention's impact (e.g. 20 percent higher profits for connected SME in comparison to comparable non-connected ones) [-];
d) the desired degree of confidence that an observed change would not have occurred by chance (the level of *statistical significance*) [+];
e) the desired degree of confidence that an actual change of the magnitude specified above will be detected (*statistical power*) [+];[6]

Only d) and e) are at the discretion of the researcher. To gauge the concrete realization of all other parameters will be difficult in most cases, though. Nevertheless, a rough power calculation conducted with approximate values will indicate how the required sample size changes if, for example, firm profits are taken as an impact indicator compared to lighting hours usage (see BLOOM 1995 for more details on sensitivity tests).

As a pragmatic alternative to power analysis, one might resort to rules of thumb: The purpose of any (quantitative) evaluation study is to compare samples of firms with each other, for example connected to non-connected firms or firms before electrification to the same firms after electrification. In order to allow for statistical analysis, as a rule of thumb, the sample size per subgroup must not fall below 30 firms, e.g. 30 connected and 30 non-connected firms. However, the number of relevant subgroups increases with the set of firm characteristics to be taken into account. For example, if the analysis furthermore distinguishes between commerce and manufacturing firms, the required sample size already increases to 120. Assuming that more firm categories have to be accounted for (regional differences, firms sizes, industries, etc.) a sample size of 200-500 often seems reasonable

[5] As a matter of course, the presentation can only be superficial at this point. For further readings on the power of surveys see also COHEN (1988).
[6] For indicators expressed as proportions (e.g. share of energy expenditures in total SME expenditures *before* the intervention) the initial or *baseline* level of the indicator additionally affects the required sample size.

and allows for the application of many statistical tools. At least for MODULE B considerations on this rule of thumb and the subgroups to account for should be provided in the inception report (see step 3b).

For MODULE C, the number of interviewed firms can be determined according to the budget. Here as well, certain differences between firms that can be important for the research questions have to be taken into account. For example, one might be interested in the (non-)use of electricity and its impacts on service firms supplying non-tradable goods and firms that are producing exportable goods as well as those producing non-exportable goods (exportable in this context refers to trade with regions beyond the intervention zone of the electrification intervention). It has to be assured, then, that at least 1-2 representatives of each subgroup – further distinguished according to their connection status – are visited.

Step 3g: Decision on sampling design

The purpose of sampling is to select firms for interviews from the total population of firms in the target region (and potentially in a control region) in a way that is governed by chance, not by the researcher's or enumerator's choice, referred to as *probability sampling*. The resulting randomness of sample selection is crucial for guaranteeing representativeness of the collected data. MODULE C is an exception by allowing as well for *purposive sampling* of firms according to specific demands or ex-ante expectations. These expectations depend on the project setup and the target region. For example, one might expect special insights on impacts in export oriented firms.

For the PRODUSE Uganda qualitative supplementary run on electricity usage in two export oriented fishing communities at Lake Victoria, to take another example, three groups were identified beforehand: voluntary non-users, "non-performers" that get connected but do not seem to benefit from the connection and "winners" that get connected and seem to be able to improve their performance. The type of firms to which MODULE C should be targeted has to be elaborated on before the survey and addressed in the inception report (step 3b). Yet, in case MODULE C is combined with another module, the researchers can decide that firms to be interviewed qualitatively are selected after the survey according to, for example, stylized firm types determined during the surveys.

For MODULE A and B some form of probability sampling has to be applied. In the ideal situation, the researchers draw a random sample from a comprehensive enumeration of firms in the whole target area. In most cases, such a list will not be available, though, only a list of villages to be electrified. Often, more than a dozen villages are electrified, so that surveying all of them is hardly an option from a logistical and budgetary point of view. The first step of sampling is therefore to select a subset of villages.[7] A random selection where the probability that a village is selected is directly linked to its population size is advisable (see e.g. IAROSSI 2007 for details). In particular for MODULE A the researcher might simply pick a subset of villages from the complete target region – either by chance or based on certain ad-hoc representativeness considerations. For example, one could choose a certain number of villages from each of different locations, in case the project intervenes in more than one region.

Per village, a certain number of firms has then to be selected – depending on the total sample size defined in step 3f. The most pragmatic approach is *simple random sampling* (within the villages): If a list of firms exists on village level, the field supervisor simply draws randomly the required number. If no such list exists, the field supervisor assigns the enumerators to different parts of the village, where the number of firms can normally be obtained from some key informant. Since SME in rural parts of developing countries are often not recognizable as such, the key informant should furthermore be consulted about the location of the individual enterprises. The first firm to be interviewed is picked by chance by the field supervisor or the enumerator. Afterwards, the enumerator visits every *n*th firm along a predefined route – with the number *n* depending on the number of firms that exist in the respective part of the village.

In brief, as long as the interviewed firms are selected randomly, a satisfactory approximation to full representativeness can be expected. Further structural sampling errors that occur in many settings can be avoided if the field research team conforms to the following two principles: (i) Cover the whole intervention area, especially in terms of centrally and remotely located firms; (ii) do not skip absent firms but revisit them later. Otherwise a

[7] In demarcation to the ideal situation of pure random sampling, this is referred to as *clustered random sampling*. Because observations from one cluster do not differ as much as observations from different clusters do, one needs a larger sample size to capture the variation between firms. The choice of the sampling scheme therefore has repercussions for the sample size determination (see step 3f and WARWICK AND LININGER 1975).

certain part of the local economy (e.g. shops that only open in the evening hours) may be excluded from the sample.

In case of the profound enterprise survey (MODULE B) the hired consultants might consider other more elaborated forms of sampling within villages, for example *stratified random sampling*. Here, firms are grouped into "strata" beforehand. Stylized firm types such as "manufacturing" and "services" are one example of strata. Geography is another logical choice for stratification, because location is likely to be correlated with a number of other variables that are of relevance for the evaluation. For a baseline study, the enterprises in a village can be stratified into "village center firms" and "more remote firms". If information on the outline of the upcoming grid is available, this may as well be used to stratify enterprises into firms located closer to the upcoming grid and those living further away.Stratified sampling assures that the two groups are adequately represented in the sample to be drawn and not – due to chance – underrepresented in a random sampling. If, for example, two in three firms in an intervention area located in the village center, two in three firms have to be surveyed there as well. For this approach it is necessary to know beforehand for each of the different "strata" the number of SME it contains.

Another option is to purposefully *oversample* firms that will more likely connect in the future in order to assure that sufficient information is obtained about them. This option is particularly relevant if the researcher worries about the risk of a low electrification rate among SME in general or among SME of a specific firm type of interest. In our example, the researcher might oversample village center firms, since one might expect that they are closer to the future power lines and therefore more likely to connect to the future grid. In the case of oversampling it is important to use weights during data analysis in order to reconstitute representativeness. Details on the implementation of the different sampling approaches and additional methods can be found in the standard literature on survey methodology (see, for example, IAROSSI 2007, MAGNANI 1997 or WARWICK AND LININGER 1975). Apart from simple random sampling, all sampling approaches should be implemented by methodologically skilled researchers.

Step 3h: Training of field work team

Interviews for MODULE C are conducted by the hired researchers themselves. For MODULE A, the field work team can consist of project staff only. If enumerators or consultants are hired they can be trained in a few hours to do the interviews, depending on the complexity of the questionnaire.

Team members hired to do the field work for MODULE B have to be trained and backstopped by a methodologically skilled researcher. During the training, the enumerators and the field supervisor have to become acquainted with the general objective of the study and the meaning and purpose of each question. Furthermore, the enumerators have to be familiarized with how to deal with non-responses, to pay attention for consistency problems and to report qualitative complementary information in comments or verbally to the field work supervisor. The training takes around 1.5 days in the "classroom" and should have an involvement of the survey team, e.g. by means of simulated interviews.

The training can be combined with a pre-test of the questionnaire, which is in this case conducted by the freshly trained enumerators under supervision of the field supervisor and the researcher. It is recommendable to contract the same enumerators for data entry afterwards. Data entry should also be taught during the training course. Pre-test and data entry training take another 2 to 3 days.

Step 4: Implementation

Step 4a: Conduct survey

In particular for MODULE A and B, a thorough logistical planning is a precondition for a successful implementation of the survey. Transport to and within the target region has to be assured. For MODULE A, one enumerator can conduct 6-7 interviews per day. The longer questionnaire in MODULE B normally makes it difficult to finalize more than 4 or 5 questionnaires per day. As a matter of course, in both cases this depends on the distance from the base camp to the survey village at the respective day and from the distance between the SMEs to be interviewed – potentially being located in more than one village.

The sampling strategy determined in Step 3f and 3g has now to be implemented in each village. In MODULE B this has to be done by the field supervisor, who assigns the enumerators to different parts of the village. The enumerators should make sure that the interviewees are the actual owners with full insights into their firm's operation – if necessary through an appointment or revisiting the firm later. In addition, it is recommended to conduct a short village level interview with, for example, the village chief to obtain an assessment on the local business environment, market access and most important infrastructure, barriers, reliability of the electricity grid, and general income sources.

Questionnaires should be checked by the field supervisor for consistency and completeness particularly after the first interviews have been completed. Potential problems and respective solutions can be discussed with the enumerators.

For MODULE C the interview length depends on the issues to be discussed with the respondent. But even if the number of questions is known, the duration is less predictable than for structured questionnaires, as spontaneous deviations from the interview guideline are possible and even desired. If enterprises state that positive or negative impacts of electrification of whatever sort exist, the researcher should – on the spot – check for other potential sources of this impact. For example, the interviewee can be simply asked if other explanations are possible for why her/his situation has improved, e.g. if the firm benefits from other development projects (in general, the comparability criteria mentioned in step 3e represent a useful starting point when trying to elicit alternative potential triggers of change).

It seems reasonable to take two hours as the maximum duration for the qualitative interview in order to not overburden the enterprise. In this case, it might also be considered to give an in-kind remuneration to the respondent to compensate for her/his loss of time. In addition, the interview might be divided and spread over the day. Thereby, the interviewer also has the occasion to observe the business at different times of the day.

Step 4b: Data entry

For MODULE A and B the entry of the collected data is a highly important step. If a proper digitalization of the questionnaire information is not assured, even the best collected data will not be useful. Therefore, much effort has to be put into preparing an easy-to-use and

trouble-free data entry template that helps to avoid data entry mistakes from the outset. In the same way, the training of staff to enter the data (preferably, this is done by the enumerators themselves, see step 3h) and backstopping the data entry (which can be done by the field supervisor) including its quality assurance are of particular importance. The best way is to supervise the entry of the first 3-4 waves of questionnaires directly and check afterwards, but still on-site, for each questionnaire whether the data is entered correctly. Once the data entry staff seems to work firmly, picking just a sample of questionnaires for quality control is sufficient.

A code sheet for additional response categories or open questions has to be provided to the data entry staff (at best after the first 3-4 waves of questionnaires have produced the most common answers) to avoid time-consuming ex-post recoding and ensure uniform usage of codes. The data can be entered in Excel and easily transferred to other statistical packages for data analysis afterwards.

For MODULE C, the data can only be entered to the extent that it is quantifiable. Depending on the number of interviewed firms, this is not always necessary. For the main body of collected information one might rather speak of "digesting" the interviews. How this is implemented depends on whether the interviews have been done by the principal researcher or by someone else. In the latter case, a systematic way of reporting the information has to be developed. This digestion step bears the potential that information gets lost and, in any case, will be time consuming – another reason for assigning the interview work directly to the researcher. At least, the staff member who conduct the interviews should be in close contact to the researchers responsible for the final report, also during the reporting phase.

Step 5: Analysis and presentation of results

Step 5a: Information and data processing and analysis

For MODULE A and B, basic data analysis can be done with Excel, which suffices to calculate frequencies, percent distributions, means, medians, and ratios. Advanced data analysis for MODULE B (regressions, difference-in-differences, matching etc.) has to be done using special statistical software packages like SPSS or STATA. These techniques can only be applied by

researchers familiar with statistics and econometrics – at best documented by a list of academic publications in the fields of impact evaluation and applied econometrics.

The applied methods should be based on the established literature on impact evaluation: RAVALLION (2008) provides a comprehensive overview of impact evaluation methods in development projects. PETERS (2009) proposes hands-on solutions in electrification projects that are feasible even with limited research budgets. Examples of applied evaluations in development projects are numerous. There are many excellent papers in the literature, but most of them have been elaborated based on surveys or data sets beyond the scope of the PU Impact M&E systems presented here. The following papers, though, are examples for methodologically proper evaluations based on limited sample sizes and can be considered as role models for methods to be applied in MODULE B: BECERRIL AND ABDULAI (2010), BECCHETTI AND COSTANTINO (2008), BENSCH, KLUVE AND PETERS (2010), KONDO ET AL. (2008), PETERS, VANCE AND HARSDORFF (2011), SCHMOOK AND VANCE (2009).

For MODULE C, the collected qualitative information has to be analysed systematically along the lines of the guiding research questions. This includes a critical assessment of who has been referred to as information sources and how to interpret the statements of the respondents.

Step 5b: Reporting

The final report of a PU Impact Evaluation effort should contain a documentation of the important steps sketched in this guideline. First, the project should be described with a focus on its *theory of change* (results chain), including activities, important steps, regional foci, objectives and intended impacts. The report has to present the study and survey implementation as well as the identification strategy. For MODULE B, there should be documentation of the extent to which the applied methods are in line with the related literature. In particular, for MODULE C, the analytical approach has to be clearly delineated in order to allow for inter-subjective verifiability.

The collected data can then be used to describe the socio-economic situation in the survey (and control) region. Only variables that are not expected to be affected by the electrification should be included in this description. The variables to be affected, that is, the impact indicators selected in Step 2b, can then be presented in an impact chapter. Sample

selection issues or other potential caveats that might distort the accuracy of the findings should be critically discussed. Not least, the researcher should try to find additional sources to cross-check findings – although it often turns out to be impractical due to lacking information on a sufficiently disaggregated level.

Analysing, understanding and digesting the collected information requires some time, which should therefore be granted to the researchers. This can reach from around 2 months in MODULE A, 3 months in MODULE C to 3-6 months in MODULE B. In particular, if advanced statistics and econometrics are to be employed, the report cannot be written in a few weeks. Note that the effective man-days to be budgeted are less. A longer period of 6 months is recommendable in order to allow for the interactive multi-stage revision process that is required to draft an understandable report on a high methodological level delivering policy-relevant results.

Step 5c: Recommendations for the project implementation

Among the different objectives of a PU Impact Evaluation System are learning effects for the project itself. Therefore, beyond the pure analysis of the data and its reporting, researchers should derive recommendations useful for the project and beyond. In the first place, of course, this concerns suggestions to improve the potential for generating positive impacts (or also to avoid negative ones). For example, the PU Impact Evaluation might reveal that regional differences in impacts exist (e.g. due to different market access or different production patterns or enterprise types). This would lead to the recommendation to focus more on certain regions, if the intention is to trigger productive use and related impacts.

A potential recommendation could as well be to modify the communication towards the public, based on which impacts could be evidenced or not. For example, in one segment the PU Impact Evaluation could document substantial benefits for the target group (e.g., households that enjoy lighting) and in another segment impacts are found to be rather modest (e.g., no substantial productive take up of electricity). The report should formulate this explicitly and recommend calibrating the communication of impacts (e.g., “Do not promise substantial productive use impacts, but highlight the social impact of the project among households.”).

Beyond the recommendations directly linked to impact results and potentials, other insights gained during the field work should be exploited in recommendations for the project managers. The field work during impact surveys always brings the researchers extremely close to the target region and its people as well as intermediate partners such as private or community operators. Experience in many projects has shown that this close work always reveals weaknesses of the project implementation as well as potentials to improve it.

5. Concluding remarks

Electrification constitutes an exogenous technology impulse that, in principle, allows micro-enterprises in rural areas to improve their productivity. In economic theory, enhancing productivity is a prerequisite for improved division of labour, which in turn enables sustainable growth. This theoretical and intuitive sequence, however, has only been sparsely investigated. Therefore, it would be very promising to generate more insights in the micro-economic development of small businesses in rural areas after electrification via PU Evaluation Systems.

Yet, the M&E budgets of electrification projects are in most cases very limited and do hardly allow for rigorous evaluation of electrification in general and PU in particular. This guideline helps to implement such research on different levels of detail and methodological rigour. Costs of the proposed approaches are manageable for most projects, in particular if a combination of the short enterprise survey (MODULE A) and the qualitative approach (MODULE C) are chosen.

The profound enterprise survey (MODULE B) might in many cases be too expensive. However, if the project particularly focuses on the development of commercial usage of electricity by, for example, complementing the electrification activities by additional business development services, including a more expensive evaluation tool is certainly reasonable. Based on a thorough research plan the projects might as well consider applying for additional research funds from independent in-house evaluation units or research programs. The present guideline helps the project managers of electrification projects to assess the needs of their project, the requirements of different PU evaluation approaches, and, eventually to implement a PU Evaluation System.

References

ADB, 2006. Impact Evaluation. Methodological and Operational Issues. Asian Development Bank. Manila, Philippines.

Becchetti, L. and M. Costantino, 2008. The Effects of Fair Trade on Affiliated Producers: An Impact Analysis on Kenyan Farmers. *World Development,* 36 (5), 823-842.

Becerril, J. and A. Abdulai, 2010. The Impact of Improved Maize Varities on Poverty in Mexico: A Propensity Score-Matching Approach. *World Development* 38 (7), 1024-1035.

Bensch, G., Kluve, J. and J. Peters, 2011. Impacts of rural Electrification in Rwanda. *Journal of Development Effectiveness*, forthcoming.

Bensch, G. and J. Peters, 2010. Socio-economic Impacts of Rural Electrification in Rwanda – An Ex-ante Assessment of GTZ Activities. *RWI Materialien* 60.

Bensch, G., Peters, J. and L. Schraml, 2010. Energy Usage and Socio-economic Conditions in Mozambique. *RWI Materialien* 56.

Bloom, H.S., 1995. Minimum Detectable Effect Size – A Simple Way to Report the Statistical Power of Experimental Designs. *Evaluation Review* 19 (5), 547-556.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences (2. Edition). Lawrence Erlbaum Associates. Hillsdale, United States.

Gertler, P. J., Martinez, J., Premand, P., Rawlings, L.B. and C. M. J. Vermeersch, 2010. Impact Evaluation in Practice. World Bank Publications. Washington, D.C., United States.

GTZ, 2007. Impact Monitoring Guide. Gesellschaft für Technische Zusammenarbeit. San Salvador, El Salvador.

M&EED Group, 2006. A Guide to Monitoring and Evaluation for Energy Projects. Monitoring and Evaluation in Energy for Development International working group. Internet: http://www.gvepinternational.org/sites/default/files/resources/MEED_Guide_final_version_english.pdf

Harsdorff, M. and J. Peters, 2010. On-Grid Rural Electrification in Benin – A Socio-economic Baseline Study on a GTZ Project. *RWI Materialien* 57.

Iarossi, G., 2007. The Power of Survey Design. A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents. The World Bank. Washington, D.C., United States.

Kondo, T., Orbeta, A., Dingcong, C. and C. Infantado, 2008. Impact of Microfinance on Rural Households in the Philippines. *Network of Networks on Impact Evaluation (NONIE) Working Paper No.* 4.

Magnani, R., 1997. Sampling Guide. Food and Nutrition Technical Assistance Project (FANTA). Washington, D.C., United States.

Neelsen, S. and J. Peters, 2011. Electricity Usage in Micro-Enterprises – Evidence from Lake Victoria, Uganda. *Energy for Sustainable Development* 15 (1), 21-31.

NONIE, 2009. Impact Evaluations and Development. NONIE Guidance on Impact Evaluation. Network of Networks on Impact Evaluation. Washington, D.C., United States.

OECD/DAC, 2002. Glossary of Key Terms in Evaluation and Results Based Management, DAC Working Party on Aid Evaluation. Organisation for Economic Co-operation and Development/ Development Assistance Committee.

Peters, J., 2009. Evaluating Rural Electrification Projects: Methodological Approaches. *Well-Being and Social Policy* 5 (2), 25-40.

Peters, J., Vance, C. and M. Harsdorff, 2011. Grid Extension in Rural Benin – Micro-manufacturers in the Electrification Trap. *World Development* 39 (5), 773-783.

Ravallion, M., 2008. Evaluating Anti-Poverty Programs. *Handbook of Development Economics*, 4, Amsterdam, Netherlands.

Schmook, B. and C. Vance, 2009. Agricultural Policy, Market Barriers, and Deforestation: The Case of Mexico's Southern Yucatán. *World Development* 37 (5), 1015-1035.

Warwick D. P. and C. A. Lininger, 1975. The Sample Survey: Theory and Practice. New York, United States.

White, R., 2002a. GEF-FAO Workshop on Productive Uses of Renewable Energy: Experience, Strategies, and Project Development. Workshop Synthesis Report, UN Food and Agriculture Organization. Rome, Italy.

White, H., 2002b. Combining Quantitative and Qualitative Approaches in Poverty Analysis. *World Development* 30 (3), 511-522.