# A European
# Ex-post Evaluation Guidebook
# for
# DSM and EE Service Programmes

***Prepared by:***


***SRC International A/S, Denmark***

*&*

*NOVEM, The Netherlands*

*Electricity Association, UK*

*Motiva, Finland*

*Norsk Enøk og Energi AS, Norway*

*Centre for Energy Conservation, Portugal*

*Elkraft System, Denmark*

*SEVEn, Czech Republic*

*Energy Savings Trust, UK*

*Wuppertal Institute, Germany*

*Regional Energy Centre of Viljandi, Estonia*


***April 2001***

# FOREWORD

This European Ex-post Evaluation Guidebook for DSM and EE Services Programmes is the combined result of two SAVE projects carried out in the period 1997-2001 (Phase I – Project No. 12488-96-12F1ED ISP DK, Phase II – Project No. XVII/4.1031/P/99-028).

The first phase of the project was completely financed by the European Commission's SAVE programme whereas Phase II was partially financed by the European Commission's SAVE programme and partially by the project team on a fifty/fifty basis.

## PHASE II PROJECT TEAM

The project team of Phase II represented a wide geographical area and various types of actors within the energy system:

- Mr. Thorkild Kristensen, SRC International A/S, Denmark (project manager)
- Mr. Preben Birr-Pedersen, SRC International A/S, Denmark (chief editor)
- Ms. Kirsten Dyhr-Mikkelsen, SRC International A/S, Denmark (technical co-ordinator)
- Mr. Harry Vreuls, Novem, The Netherlands
- Mr. Derek Baggs, Electricity Association, UK
- Ms. Ulla Vuorio, MOTIVA/Finnbarents, Finland
- Ms. Liv Randi Lindseth, Norsk Enøk & Energi AS, Norway
- Mr. Carlos Nascimento, Centre for Energy Conservation, Portugal
- Mr. Mikael Togeby, Elkraft System, Denmark
- Mr. Martin Dasek, SEVEn, Czech Republic
- Mr. Andrew Amato, Energy Saving Trust, UK
- Mr. Stefan Thomas, Wuppertal Institute, Germany
- Mr. Kalle Jöks, Regional Energy Centre of Viljandi, Estonia.

The project team represented a variety of organisations and countries, something, which will help ensure that knowledge of the existence of this guidebook is widely spread. For details on how to contact these organisations, please consult Appendix A of this report.

Although non-project team members provided valuable comments, the full responsibility for the outcome of the project resides with the project team.

## PHASE I

Phase I was organised by JRC, Italy with SRC International, Denmark as technical co-ordinator.

The objective of Phase I was to establish an overview of the existing evaluation practices used, available methods, and what issues the new methodology should incorporate.

Phase I consisted of the following main tasks:

- Bibliographic research of existing methodologies evaluating DSM and EE services programmes and preparation of summaries of selected methodologies.
- Translation of selected parts of the Swedish Evaluation Guidebook, prepared by NUTEK in 1993, relevant to other European energy markets.
- Preparation of a comprehensive draft document on general ex-post evaluation methodology based on the bibliographic research and the Swedish Evaluation Guidebook.
- Preparation of a comprehensive draft standard reporting format linked to the developed ex-post evaluation methodology The format was to be used for reporting on DSM and EE programmes in a consistent and logical manner which allows comparison of programme and evaluation results (at intra-company, regional, national, and international level).
- Identification of organisations to participate in Phase II and a selection of evaluation experts who reviewed the findings on a continuous basis throughout the project period.

The outcome of Phase I was published January 1998. This publication supersedes the Phase I results.

## PHASE II

The objective of Phase II was to test the drafted methodology in various environments in terms of programme objective, implementation method, and market structure. Participating organisations were requested to test the drafted methodology on a specific DSM or EE services programme. The programme could be a programme implemented by the participating organisation itself or others. The project team provided support and guidance to the participating organisations as they applied the evaluation methodology to their own programmes. The purpose was to provide hands-on experience to the participating organisations and obtain valuable feedback on the practicality of the draft guidebook. In this way, the guidebook has benefited from real-world experience of organisations carrying out evaluation for this final document to be as practical and useful as possible.

Phase II thus included:

- Testing of the draft European ex-post methodology using the comprehensive draft report prepared in Phase I and the standard reporting format.
- Reporting on the adequacy of the draft European ex-post evaluation methodology and the standard reporting format.
- Preparation of illustrative case examples. The mix of case studies was composed in such a way as to reflect particularly interesting aspects, which have not been addressed to a great detail by previous reports.

- Review and modification of draft methodology and preparation of the final project document on the European Ex-post Evaluation Guidebook.
- Presentation of the European Ex-post Evaluation Guidebook at the ECEEE Summer Study 2001 and other EE and evaluation fora to allow dissemination of the experience gained and the guidebook itself.

The present report constitutes the final European Ex-post Evaluation Guidebook.

# CONTENTS

# 1 INTRODUCTION

## 1.1 WHY A EUROPEAN EX-POST EVALUATION GUIDEBOOK?

Numerous demand-side management (DSM) and energy efficiency (EE) services programmes have been carried out over the last two decades; some supported by the EU SAVE programme.

However, many utilities and energy agencies have carried out costly campaigns and made large investments in EE programmes with only limited assessment of their impact. In other words, without a clear idea of:

- How much energy is truly being saved;
- Where the saving is taking place and why;
- To which extent greenhouse gas emissions are being reduced;
- Whether programmes are being operated cost-effectively.

The introduction of competitive energy markets and limited resources forms a growing pressure to prove impacts and justify costs.

Some hope to use EE programmes and projects to persuade customers to remain with the local utility or to gain market shares and need methods for evaluating the cost-effectiveness of their efforts in the competitive market environment.

Energy service companies (ESCOs) – both utility subsidiaries and independent ESCOs – need evaluation expertise, to assure their customers that promised energy savings have indeed occurred and to limit their investment risk.

Finally, the growing pressure on local energy agencies to compete with private businesses for the EE activities creates a need for proof of impact and justification of costs, also within the public domain.

Development of an evaluation guidebook is thus a logical continuation of national and international efforts to meet $CO_2$ reduction targets and limit energy consumption.

## 1.2 GUIDEBOOK OBJECTIVES

The overall objective of the project was to disseminate evaluation theory and thus indirectly help reduce the overall $CO_2$ emissions and improve energy efficiency. Evaluation of implemented DSM and EE services programmes provides useful information and allows optimisation of the programme cost-efficiency. Pursuit of more cost-efficient programmes means increased value for money. If money can be saved or better employed this could help lead to a greater reduction of $CO_2$ emissions. Implementation of DSM and EE services programmes without serious evaluation of the achieved impact is increasingly hard to justify.

The immediate objective of this guidebook is to present a European ex-post evaluation methodology for DSM and EE services programmes. In spite of it being an ex-post methodology, it is also envisioned as a planning instrument for new DSM and EE programmes and many of the elements presented here may also be used for other types of evaluation.

Demand-side management (DSM) is an activity designed to influence the energy demand of consumers. It was developed to complement supply-side management. The term “DSM” originally referred to activities carried out by energy utilities or governments in a monopoly market to reduce energy production and delivery costs and energy consumption. The term “EE services” typically refers to the services required to specifically ensuring that conversion of energy at the customer’s premises is as efficient as possible, given the technical and financial constraints. EE services can be carried out with the aim to increase income of the service provider as opposed to reducing energy production and delivery costs. EE services may for example be part of a customer retention programme or an ESCO project.

In the guidebook no clear distinction is made between DSM and EE services. Together they cover all attempts at manipulating the demand-side consumption and improve demand-side energy efficiency. Instead, a distinction between programme types is used (targeted information, market transformation, transmission & distribution, load management, customer retention, and ESCO).

No clear distinction is made between the terms “programme” and “project”; the term “programme” is used as a common term in most of the guidebook text.

Furthermore, the term “programme” is used in its widest meaning. It includes very small programmes (by some referred to as projects) to very large programmes.

Also, no terminological distinction is made between large-scale (national) programmes, which consists of several individual programmes/projects and a single programme. The main difference between the two, in terms of evaluation, is that the large-scale national programme may require evaluation of the priority given to the individual programmes (i.e., should the resources be distributed differently) and their influence on one another. The question of whether some of the existing programmes/projects should be excluded and new programmes/projects included in the national programme is a question which is also relevant for independent programmes (i.e., would another programme have been more cost-effective; should we choose another programme for the next period; etc.).

**The guidebook aims to provide arguments, questions, and examples on EE evaluation.** Sound evaluation is to a certain extent tailor-made and requires innovative and creative thinking. The guidebook provides examples that illustrate the methodologies that can be used.

The guidebook is a hands-on document, addressing evaluation needs in both captive and competitive energy markets. It describes available evaluation methods, their possible application, and a detailed, step-by-step description of **how to plan and implement** them.

The user of the guidebook should be able to plan and implement an evaluation with the guidebook as his/her main inspiration. The target readers/users are people who carry out actual evaluation and not theoretical experts. Given that the readers will have varying skills in the field of evaluation, some readers will find the guidebook more useful than others. Whether novice or a more experienced evaluator, it is, however, recommended that you read Section 2.5 and Chapter 3.

It is the hope of the project team that this guidebook may help increase the number of programmes being evaluated as well as the quality of the evaluations.

## 1.3 RECOMMENDATIONS

Recommendations from the authors and main points made in the guidebook are:

- **Do carry out evaluation**. The gained information from a good evaluation is vital for improvement of programme activities. **Use the results of the evaluation actively**.
- Take **care in formulating the objectives** of your evaluation. This facilitates the task, increases the credibility of results, and limits costs.
- **Start planning the evaluation as early as possible** in relation to programme design. A lot can be gained from planning the ex-post programme evaluation even before the programme is carried out. It is of course not necessary and, alas, far from normal practise in today's world. However, a significant amount of work can be saved and more reliable data can often be gathered at a lower cost by early planning.
- Arrange **good communication** between the various parties involved in and affected by the evaluation and be aware of the possible consequences of the evaluation results. Care should be taken to present the evaluation results so that they are immediately useful to the intended users.

## 1.4 GUIDEBOOK STRUCTURE

The challenge in developing a guidebook is to get the right balance between providing an overview and providing detailed "how to do it" information.

One of the main concerns in the preparation of this guidebook has been to **avoid duplication** of information, which can already be found in other texts while providing the reader with a document detailed enough to function as a guidebook. The guidebook thus concentrates on drawing attention to the issues and planning steps, which the evaluator must address to achieve a successful evaluation.

Evaluation starts with determining the framework for the evaluation, i.e., establishing why the EE programme was implemented and why there is an interest in evaluating the programme (**Chapter 2**). Although evaluation needs are not necessarily directly linked to the programme objectives, the focus of interest can vary depending on programme type. For example are the interests associated with a customer retention programme different from that associated with a market transformation programme, which could lead to a difference in evaluation objective. Attention has been given to clarifying this link.

The choice of evaluation method will depend on a multitude of factors including experience (staff), time, available budget, users of the evaluation results, and programme specific circumstances.

**Chapter 3** presents the organisational planning aspects of the evaluation effort. It offers a step-by-step approach to evaluation spanning from definition of objectives to selecting evaluation strategy. It also briefly touches upon the issue of evaluation budget.

General guidelines for selecting an energy impact evaluation strategy are provided in **Chapter 4** while **Chapter 5** outlines specific evaluation concepts including how to estimate the baseline development, i.e., the development which could have been expected provided that the programme had not been implemented. **Chapter 6** demonstrates evaluation strategies for specific programme types. The programmes categories used are targeted information, market transformation, transmission & distribution, load management, customer retention, and ESCO.

Apart from evaluation of the energy impacts of a particular programme, evaluation of the project process can provide valuable information to the programme planner, based upon which project cost-effectiveness can be improved. Related key issues are presented in **Chapter 7**. The techniques presented here may also be relevant for market impact evaluations.

**Chapter 8** gives an introduction to how to apply evaluation results.

An **index** has been included at the back of the guidebook to facilitate easy reference to the parts of the guidebook of particular interest to the individual reader.

The guidebook methodology has been tested by ten European experts on specific DSM and EE services programmes. The evaluation cases cover a variety of target groups, technologies and approaches and have been carried out by both experts and novices in utilities, research organisations and governmental organisations. **Appendix A** contains a short presentation of each of the evaluation case studies. The descriptions have been used to form illustrative examples, which have been placed in the main part of the guidebook ("grey boxes").

**Appendix B** contains information on further reading and lists relevant web sites and conferences, as well as provides summaries of selected evaluation methodology literature, which supplement the guidebook methodology. Bibliographic research was carried out to benefit optimally from existing methodologies and to avoid unnecessary repetition and overlapping between these and the European ex-post evaluation methodology. Summaries of the most relevant documents were prepared to provide the reader with easy access to this information. Furthermore, the summaries are intended used actively as a supplement to the main report and the reader is encouraged to exploit the opportunity. An overview of the summaries is presented in Exhibit 1-1.

**Appendix C** contains a summary of the work on a standard reporting format**{** XE "standard reporting format" **}** for DSM and EE services programmes and their evaluation. Also presented there are four examples of formats.

An important outcome of the project, which is not visible in the guidebook, is the **network** created between the involved experts/organisations – a network that will continue to operate in the future. Sparring with colleagues and experienced evaluators is an effective way to promote successful evaluation.

*Exhibit 1-1: Overview of bibliographic summaries.*

| # | Title | Contents |
|---|---|---|
| 1 | Realistic Evaluation | New trend in evaluation. The evaluation should not focus alone on average/aggregated impact since great information exists in the detail. |
| 2 | Guidelines for Defining and Documenting Data on Costs of Possible Environmental Protection Measures | These guidelines establish a common framework and vocabulary for documenting and using data on costs. |
| 3 | Evaluation, Verification, and Performance Measurement of Energy Efficiency Programmes | Standard evaluation methods - process, market, and impact evaluations including various impact evaluation methods to address various programme and technology types. |
| 4 | Evaluating Energy-Efficiency Programmes in a Restructured Industry Environment | Methods aimed at evaluation in restructured markets including critical review of standard evaluation methods and extensive discussion of information programme and market transformation programme evaluations. |
| 5 | Market Transformation in a Changing Utility Environment | Methodological issues regarding market transformation programme evaluation. |
| 6 | European B/C Analysis Methodology - A Guidebook for B/C Evaluation of DSM and Energy Efficiency Services Programmes. | Guidebook for performing benefit/cost analyses of DSM and EE services programme including identification of benefit and cost trade-offs between the various perspectives in designing and evaluating DSM and EE services programmes |
| 7 | International Performance Measurement and Verification Protocol | Standard methods for measuring and verifying savings for energy services projects. |
| 8 | Evaluating Market Transformation Initiatives: Issues, Challenges, and State of the Art | Establishment of the purpose of evaluation efforts. Key issues and challenges in evaluation of market transformation programmes. |

# 2 PURPOSE OF EVALUATION

## 2.1 WHY PROMOTE ENERGY EFFICIENCY?

EE activities can be classified into two basic types: **Public policy** EE sponsored or directed by governmental and other public sector organisations, and **business-related** EE sponsored or directed by energy firms. There are reasons to implement each type of activity, both in captive energy markets and in more competitive ones.

### 2.1.1 RATIONALE FOR PUBLIC POLICY EE ACTIVITY

Many EU countries have developed public policy strategies to implement DSM and EE programmes. These programmes have been implemented by governments and, to a lesser extent, utilities. However, most of these programmes have been developed in an era of government-owned or -regulated, captive, monopoly businesses.

Restructuring the electricity supply industry changes the relationship between customers, utilities and regulators. In a monopoly environment, the government can rely on an implicit social contract between government and utilities as the policy basis for using utilities to implement social policy. As the monopoly is withdrawn, altered or changed, the basis for using utilities weakens and may even vanish altogether. However, rationale for public policy EE activities remains.

A publication produced by the International Energy Agency summarised the effects of energy restructuring on the rationale for public policy EE activity, as shown below.[1]

Ample literature and public policy debate has shown the weaknesses of monopolistic electricity sector in relation to achieving an economically justified level of energy efficiency. Proponents of intervention have identified a number of **market failures** that cause consumers not to choose a level of energy efficiency that appears to be economically justified. These market failures include:

- Pay-back gap;
- Prices differ from marginal cost;
- Risk sheltering of the utility;
- Externalities
- Lack of information and high transactions costs;
- Disconnected decision-maker;

[1] Review of Existing Mechanisms for Promoting DSM and Energy Efficiency in New Electricity Business Environments, *The IEA DSM Implementing Agreement, A Report of Task IV — Development of Improved Methods for Integrating Demand-Side Options into Resource Planning and Government Policy (Subtask IV/6), Final Report - October 1996, pp. 33-37.*

- Lobby effect; and
- Consumer capital constraints.

These are a selection of the range of market failures that have been cited as a rationale for intervening in energy markets to encourage higher levels of energy efficiency. For further readings, please refer to the report "Methodologies for Evaluating Energy Efficiency Programmes in Central and Eastern Europe", IEA, October 1996.

A number of the market failures identified above are directly linked to the existence of a monopoly in electricity supply and the type of regulation that has been applied to limit monopoly electricity utilities' market power. With restructuring of the electricity sector to encourage (1) competition in generation and (2) full customer choice of their electric supplier, some of these market failures are either reduced or eliminated while others remain, and maybe new occur.

Therefore, a rationale to encourage DSM and energy efficiency still exists in a competitive market, though the nature of the market failures will have altered. Determining whether or not to intervene in the electricity market to encourage DSM and energy efficiency will undoubtedly be a political decision. In developing an effective policy for intervention, policy makers would be well-advised to be as specific as possible in identifying the market failures that the intervention is intended to influence, and to prioritise (to the extent possible) the competing objectives of intervention.

### 2.1.2 RATIONALE FOR BUSINESS-BASED EE ACTIVITY

Three primary entities have rationale for implementing business-based EE activity:

- **Retail energy marketers** – The motive will be customer retention or profit generation. While some suppliers may seek to be the absolute lowest-cost provider, many may offer a competitive price but try to compete on some other basis. Many analysts believe that an effective strategy is to combine provision of the energy commodity with other products and services, which might, as a package, provide specific customers with higher value overall and therefore help in retaining existing customers and attracting new ones.

- **Electricity distribution companies** (in competitive energy markets) – In addition to a possible role in some types of public policy EE initiatives, electricity distribution companies may also have a self interest in implementing geographically targeted DSM/EE activities that can delay or eliminate the need for costly transmission and distribution (T&D**{** XE "T&D" **}**) system upgrades, e.g., peak load reductions.

- **Energy services companies** (ESCOs) in both regulated and competitive energy markets – These have EE activities as their primary business. The services may include efficiency upgrades of facilities and energy-using equipment; design, construction and/or turnkey implementation of efficient energy-using equipment; and performance contracting. The primary rationale is a fee for service, to generate profit for the ESCO.

## 2.2 REASONS FOR EVALUATION

Evaluation assesses programme effectiveness (are immediate objectives met?) and/or programme efficiency (could the objectives have been met with a lower use of resources?).

There are three main reasons for evaluating EE programmes:

- To estimate programme **impacts**, including:
  - Energy demand related impacts (e.g., energy use, capacity demand, greenhouse gas emissions, or market barriers);
  - Business-related impacts (e.g., impact on customer retention rates, profit margins, or overall profitability);
  - Market reactions to the programme (e.g., profiles of market segments participating and not participating, effects on equipment manufacturers, suppliers and market channels);
  - Explanations behind programme impact estimates (i.e., how and why the programme's impacts were what they were).
- To determine how the **process{** XE "process" **}** of the programmes could be improved, including:
  - Efficiency of programme procedures, programme outreach and information processing;
  - Methods for streamlining the programme and improving cost-effectiveness;
  - Effectiveness of marketing strategies and promotional materials;
  - Participant satisfaction with the programme (some analysts consider this issue relevant in both process and impact evaluation).
- To meet **contractual requirements** – Some energy services companies engage in performance contracting work, in which at least some portion of payment for services provided is based on the performance of the installed energy efficient equipment. Evaluation (monitoring and verification) requirements are typically written into such contracts, and specify the item to be measured, the way in which the measurement will occur, the duration of the measurement, and the frequency of the measurement.

Historically, most evaluations have had as their central objective the assessment of the reduction of energy use or $CO_2$ emissions**{** XE "$CO_2$ emissions" **}**. Consequently, impact evaluations have received most of the attention. For EE projects implemented by energy service companies in the private market, this is the sole objective.

As programme objectives have shifted to transforming markets and overcoming market barriers to energy efficiency, the focus of the impact evaluation has shifted to a more approximate estimate of energy use and $CO_2$ impacts and a more detailed estimate of market-related impacts, supported by market evaluation data on key market indicators (see Section 3.3).

When programme sponsors have been concerned about cost-effectiveness and public perceptions, or when programme implementers have had to prove to regulators or company management that programme funds were spent prudently, process evaluations have also tended to play a key role. Process evaluations typically trace the flow of the programme from programme design to measure implementation, examining programme marketing, customer contact, participation processing, and programme monitoring.

## 2.3 RISKS AND BENEFITS

Decision-makers have two clear options with regard to evaluating their programmes:

- **Not to evaluate** the programme. The risks are that:
  - The programme may be implemented inefficiently because previous experience is not used to the full extent.
  - Financial and human resources may be wasted.
  - Individuals and organisations may make decisions based on incorrect information (e.g., believing that a programme is saving energy when it is not, believing that it is saving much more than it is, or believing that it is cost-effective).
  - The programme may cause unexpected, and possibly undesirable, effects in the market place.
  - Others may not accept claims that energy has been saved, or claims that energy has been saved in the amount reported.
- **To evaluate** the programme. If this is the choice, further decisions must be made as to the level of detail for the evaluation as well as the methodology to be used. Implications of this choice are as follows:
  - Resources will be spent.
  - Human resources will be tied up.
  - Decision-makers and programme managers will obtain a better and more complete understanding of the effectiveness and efficiency of the programme.

Furthermore, evaluation may be required by law or due to contractual obligations (e.g., the Danish Integrated Resource Planning Law).

## 2.4 ANSWERING PRACTICAL QUESTIONS

In practical terms, evaluation can address practical questions that decision-makers need answers to, such as the following questions.

### 2.4.1 IMPACT QUESTIONS

**Question:** How can I **prove** (to the government, citizens, the media, other countries, my manager, my customer) that this programme or project is having significant impacts (on energy use, $CO_2$ emissions, customer retention, profit margins)?

**Question:** How **accurate** are the programme's initial assumptions regarding specific impact parameters?

In designing EE programmes, values are estimated for a range of factors, and those values directly affect the programme's cost-effectiveness. If the evaluation shows that the values for these factors are really much higher or much lower than assumed in the design phase and that the programme is not cost-effective (from the point of view of the programme sponsor, the programme participant, or society) then funds can be shifted to other programme options or programmes. An evaluation can reveal key data regarding components such as:

- Number of hours the targeted piece of equipment is used daily — Typically, the more hours the targeted equipment is used, the more energy savings accrue.
- Level of use of the equipment — Especially for commercial/industrial energy-using equipment, sometimes equipment can be used at less than full capacity (oversized in anticipation of future growth or operated in parallel with similar equipment as a reliability measure), resulting is lower than expected energy savings.
- Actual efficiency of equipment being replaced and of the more efficient equipment promoted by the programme — Nameplate energy load ratings of equipment can be inaccurate, resulting in lower- or higher-than-expected energy savings.
- Percentage of programme participants who have removed, disconnected or are otherwise not using the programme measure.

**Question:** What changes in energy use are occurring specifically as a result of the programme, compared to changes that would have occurred without the programme (net impact)?

**Question:** Has the programme caused **reduced or increased energy use** among participants with regard to other equipment or behaviours? Has it affected how even those who did not participate use their energy?

**Question:** Are some participants **trading energy bill savings** for other benefits, such as an increased winter thermostat setting or extended work shifts to achieve greater production at the same energy cost? What effect is this behavioural change having on the programme's overall objectives (e.g., reducing greenhouse gas emissions versus positive public relations)?

**Question:** How is the programme **affecting markets**? For example:

- Are new retailers/vendors of the equipment promoted by the programme entering the market?
- Is the number of product lines (models) of the targeted product increasing?
- Is product availability increasing?
- What effect is the programme having on competition between similar efficient products?
- Is market share for efficient equipment increasing?
- Are standard stocking practices and ordering procedures changing to facilitate the purchase of the efficient equipment?
- Is the cost of similar products decreasing, due to elevated competition (e.g., the cost of all compact fluorescent lamps, due to a programme promoting certain models)?
- Is the subsidisation of efficient products and services having a harmful effect on certain manufacturers or energy service companies, by devaluing the product/service they are

offering? Is the programme having a positive or negative effect on the energy service industry in general?

**Question:** Is the programme **still needed**? For example:

- Are there signs that the equipment or practice being promoted is becoming standard in the market place? Is demand for it increasing? Do customers expect it? Is less efficient equipment difficult to find?

- What is the likelihood that codes or standards requiring the targeted equipment or practice will be enacted in legislation (possibly because the programme demonstrated the practical feasibility or benefits of the measure)?

- Are the programme costs much higher in value than the programme benefits? Could the public's or the company's resources yield more value if diverted into a different type of efficiency effort, or a different effort not related to energy efficiency?

**Question:** What **motivated** the target group to participate and implement energy efficiency (functional benefit, economic benefit, ecological benefit, aesthetic benefit, social benefit) (relates to the issue of self-selection, see Section 4.2) ?

### 2.4.2 PROCESS QUESTIONS

**Question:** To what extent are programme funds being **wasted** on activities to persuade individuals/businesses to take actions they will take even if no funds are spent?

- Exactly how is the programme marketed, who is involved and when?

- How is the participation process supposed to flow, from customer contact through measure implementation (or incentive payment, final inspection, if appropriate)?

**Question:** Can the same or greater impacts be achieved using a different **programme structure/design** with lower costs? For example:

- Why are expected impacts not occurring?

- Are undesirable effects from factors/forces external to the programme negatively affecting the programme? How? Can they be reduced in some way?

- Could the programme take advantage of synergies or economies of scale through joint efforts with other entities?

- Is programme marketing most influential with low-impact consumers?

- Would a different marketing message or different marketing channel result in a greater percentage of high-impact consumers[2]?

*[2] Low-impact participants are those participants for whom the equipment or behaviour targeted by the programme results in less than average per-participant energy reduction. Perhaps they use energy-using equipment less often, for shorter periods of time, during off-peak times, have heating thermostats at lower settings, etc. Conversely, high-impact participants' use of the equipment or behaviour targeted by the programme results in more than average per-participant energy reduction.*

- Is some portion of the marketing or promotional effort ineffective? Could it be terminated, resulting in a lower cost per participant and lower cost per unit of impact?
- Would participants make the same efficiency improvements …
  - if they had access to low-interest loans rather than incentive payments?
  - if they could pay for the efficient product in instalments as part of their electric bill instead of receiving any financial incentive at all
  - if someone would mail a full-price efficient product to them rather than offering a rebate on the product when purchased from retailers?
  - if they could receive information from a source they trusted that the product really worked as claimed, rather than receiving a rebate to purchase the product?
- Is there unnecessary redundancy in the processing of programme records? How many individuals in how many locations handle programme forms? Should this activity be more centralised or more dispersed?

**Question:** Are programme **funds being spent wisely**? For example:

- Were programme implementation contractors selected based on personal relationships instead of best value?
- Are programme marketers given an incentive based on number of participants rather than overall energy savings resulting from the participants?
- Does in-house staff do significant field implementation work? Does this drive up cost? Would outside contractors be more cost-effective or provide a higher quality service?
- (Especially for programmes of utilities operating in captive markets) Is the utility responsible administering the programme limiting the range of efficiency improvements possible, due to mixed messages from management about saving energy versus selling energy? Does programme implementation result in only certain types of equipment being implemented (e.g., electric), when the programme was supposed to be designed to be fuel-blind?

### 2.4.3 CONTRACT REQUIREMENT QUESTIONS

**Question:** Is the agreed level of performance reached?

**Question:** Can the agreed level of comfort be reached at a lower cost?

**Question:** Is the client still satisfied with the energy service contract or could the client be interested in an alternative or additional service?

**Question:** Were the chosen contract indicators appropriate?

**Question:** What were the advantages and disadvantages of using a contract approach to EE activity?

The answers to questions such as these contribute to answering the **higher level question**, "Should we be implementing this programme or this type of programme?" They also address accountability concerns, both by providing information on how effective the programme actually is and by communicating to those involved in implementing the programme that management/the programme sponsor is concerned about cost-effectiveness and prudent use of funds.

# 2.5 WHAT IMPACT VERSUS WHY THIS IMPACT

Typically evaluations focus on what impact was achieved due to a certain programme rather than investigating what caused this impact. However, understanding of the latter is very important for improving programme effectiveness and efficiency as well as for reproducing a successful programme.

Much too often the theory behind an evaluation is implicit – it is not published and it may not exist. Without any theory, the results from an evaluation are somewhat like a "black box"*: An energy audit was done in 50 companies and X MWh was saved*. Why the X MWh were saved, or what was important for the result is not known. The evaluator might have an idea, e.g., that it was the focus on profitable projects that made the audits successful, but in reality the mechanisms behind the impact are not known or are not investigated. Problems occur when a programme is repeated in another context, e.g. in another location, with other staff, or with a reduced scope. Without theory it is difficult to know whether the changes are important for the success or not.

We wish to point to three things related to impact evaluation:

- Cause-impact relationship and programme/problem mechanisms;
- Context and timing dependence;
- Importance of variations.

## 2.5.1 REALISTIC EVALUATION THEORY

According to realistic evaluation theory:

> "Evaluators need to focus on how the causal mechanisms, which generate social and behavioural problems, are removed or countered through the alternative causal mechanisms introduced in a social programme. Realist evaluators seek to understand why a programme works through an understanding of the action of mechanisms. Mechanisms refer to the choices and capacities, which lead to regular patterns of social behaviour. Causal mechanisms are at work in generating those patterns of behaviour, which are deemed social problems and which are the rationale for a programme. Programmes are often prolonged social encounters and even the simplest initiative will offer subjects considerable compass for decision making. A key aspect of evaluation research design is thus to anticipate the diversity of potential programme mechanisms involved and a key analytic task is to discover whether they have disabled or circumvented the mechanisms responsible for the original problem." (Pawson & Tilley, 1997, page 216)

The term "problem" refers to the unwanted state of things, which the programme targets (e.g. unrealised cost-effective energy saving potentials in industry).

The great challenge of impact evaluations is (according to the realistic evaluation theory) to determine the **cause-impact relationship**. One task is to document that a change has taken place in the size of the problem or the character of the problem. Another and far more difficult task is to document that the change (or part of the change) happened due to a certain effort (i.e., the programme).

Often, it is the impact of a specific activity carried out at a given time and place, which we want to analyse. However, the outcome of such an analysis depends among other things on the **context{** XE "context" **}** and the **timing** of the evaluation. When repeating the activity, a larger or smaller amount of conditions will invariably have changed. The activity may for example be repeated in a different location, the implementers might have different qualifications or the extent and terms of the financial support might be different. If important parameters are modified, then it is natural to assume that it will affect the programme impact. It is therefore vital in a repetition or transfer of a programme to know, which parameters, i.e., **mechanisms{** XE "mechanisms" **}** influence the problem. The mechanisms describe the personal arguments for a given behaviour. Significant mechanisms should be included in the evaluation and be covered in the data collection. Furthermore, trustworthy arguments for the lack of significance of the excluded mechanisms must be formed.

The mechanisms of a problem or a programme can be described explicitly in a theory. The theories may describe the programme or they may cover the problem, which the programme is trying to address. **Programme theory{** XE "programme theory" **}** for a labelling system for the energy consumption of cars describes how this information influences the buyers. The **problem theory{** XE "problem theory" **}** describes what influences the buyers to choose an energy efficient car or not. Combining the theories with information about the context and timing of the programme, it becomes possible to assess why the impact occurred.

The programme impact most likely varies depending on the buyer's situation (and the general context and timing). Not all car buyers react in the same way to energy labelling. **What works for whom in which context** is more important than average impacts if you wish to learn from your activities. An evaluation that only establishes the total programme impact or average impact will therefore miss out on important information i.e., in which circumstances the programme is effective and in which it is not. **Variations{** XE "variations" **}** and context contain useful information, which can help improve the programme/problem theory and programme design.

A more extensive example of programme mechanisms and problem mechanisms is given in the following based on the Danish Campaign for Lower Clothes Washing Temperatures (see also Appendix A for more detail on the evaluation). An assessment of the relative importance of the different mechanisms and their interdependence would lead to the formulation of a complete programme theory and a problem theory for the clothes-washing programme. A description of the relative importance and the interdependence of the following mechanisms have, however, not been included in the following.

The mechanisms marked "time related" are particularly sensitive to the timing of the intervention.

**Problem:**

High electricity consumption for washing in residential households constitutes 4.5% of total household consumption. The 90°C clothes-washes make up 15% of the consumption for clothes washing, which is considered too high.

**Applied Method:**

Dissemination of information.

**Message:**

Clean washing is possible/best at 60ºC (as opposed to 90ºC).

**Indirect Message:**

It takes energy to wash clothes and this has an impact on the environment.

**Programme Mechanisms (How does the message influence the washer):**

- **Conscious acting** – The washer starts thinking about how she/he washes clothes and why.
- **Individual decision-making vs. tradition** – The washer consciously decides based on the present level of information available how to wash clothes instead of doing what her/his parent(s) did.
- **Social value** – Clothes washing becomes an area of social attention e.g. public media attention; men may show an interest in how the clothes are washed; clothes-washing becomes an acceptable subject of discussion at social gathering; etc.
- **Washing of clothes is linked to environmental concern** – Environmental protection is introduced as an issue in relation to clothes-washing (indirectly affected through the associated the energy use) as opposed to only hygiene and dirt/stains.
- **Peace of mind** – Washers get satisfaction knowing that fabric care (i.e., lower washing temperatures) is not opposed to hygiene.
- **Snowball effect** – The campaign opens up for taking in information already available for example on detergent packaging regarding adjusting the amount of detergent to the water hardness and the type of wash.
- **Male interference** – Assuming that washing of clothes is still very much a female task (traditional gender roles), it might be seen as male interference (i.e., programme provider = male) in a female territory to "dictate correct procedures".
- **Attitude to programme approach** – The washer's reaction to the programme message depends on the approach used.
- **Attitude to programme provider** – The washer's reaction to the programme message depends on the washer's attitude to the programme provider (e.g. is the provider perceived as trustworthy).
- Et cetera.

**Problem Mechanisms (Who washes at 90ºC and why):**

- **Number of households with washing machines** – Determines the size of the potential target group.
- **Number of persons per household** – Can the household fill the machine completely at each wash if they want to?
- **Weather and climate** (time related) – Weather determines how many clothes are worn, what clothes are worn and how often they are changed. In Northern Europe more white

clothes are worn during summer (and especially during hot summers) and more cotton/linen are used which lead to a higher usage of 90ºC washes to remove sweat and stains. Colder climates require larger pieces of garments and/or a greater number of garments, which may lead to more washing depending on the need for symbolic cleaning.

- **Fashion** (time related) – Decisive factors are fabrics (not all fabrics can be washed at 90ºC), some cannot be washed at all), cuts (a tight fit will invariably lead to a greater need for washing), and colours (not all colours are resistant to 90ºC washes). These factors may work for as well as against 90ºC washes.

- **Distribution of professions** – Some types of work is dirtier than other types. Some work requires clothes, which is dry-cleaned and not washed. Some work places handle the washing of work clothes for their employees.

- **Spare time activities** – The tendency in Western countries is that each spare time activity requires its particular outfit. With many activities per household this increases the wash load especially when combined with a need for symbolic cleaning.

- **Age distribution of population** – The need for clothes washing depends on the age of the wearers, i.e., the persons within the individual household.

- **Symbolic cleaning** – The frequency of clothes washing is high since one-day wear of clothes is the social standard; you should not be seen wearing the same clothes two days in a row (due to a hysteric attitude towards human smell).

- **Age distribution of washers** – Young people of today are trained to collect and use information from many sources while older generations are more bound by tradition i.e., doing what their parents recommended/did (use information provided by their parents irrespective of the change in context). Furthermore, the younger generation may be more prone to behavioural change as such (i.e., their worldview is not cemented).

- **Sex of washer** – There may be differences in the way women and men reason and act in relation to washing of clothes and adopting a message as the one in question.

- **Key decision-maker** – A domestic servant washing the clothes of his/her employer might not have freedom of choice regarding temperature or even have an interest in the matter.

- **Economic wealth of the households with washing machines** – A wealthy household is less likely to care about potential energy bill savings.

- **Awareness and sense of responsibility in sustainable development** (time related) – The number of people, who wish to act upon their environmental conscience, can be influenced by other issues on the political agenda at the time of the programme. The presence of other EE campaigns may for example indirectly increase the impact of the campaign while a large political crisis may crowd the public attention.

- **Personal/societal history** – A country exposed to an energy crisis could seriously tune people into energy efficiency including EE behaviour when washing clothes. The opposite is also possible: A generation which in the past became exposed to a disease epidemic, which could be limited/contained by washing clothes at 90ºC, is more likely to uphold "old" clothes washing habits.

- **Design of the electricity bill** – Visibility of the link between washing of clothes and the size of the electricity bill influences the impact of the message.

- **Price of electricity** – A low impact on the overall household budget can mean that the household is less prone to EE behaviour.

- **Contact with dissemination media** – The choice of media for dissemination of the programme message will determine who comes into contact with the message as well as how it is received.

- Et cetera.

### 2.5.2 REQUIRED EFFORT FOR PROGRAMME THEORY DEVELOPMENT

An in-depth analysis of the cause-impact relationship puts great demands on theory and analysis methods. A precise proof of the influence of both significant and less significant parameters also often requires a large data material. This may result in impossible or unreasonably expensive evaluations. Such evaluations cannot be carried out on discount budgets. Successful impact evaluations often cost a lot of money and require years and not weeks. In-depth impact evaluations are similar to research projects. Especially small programmes cannot justify these kind of intensive evaluations. A lot can, however, be gained by focusing the evaluation properly and documenting the programme context, timing, and variations clearly and explicitly for future reference. Such documentation is very useful when developing a new or revising an existing programme since it identifies the characteristics of a successful programme design and marketing strategy. It provides an understanding of why a programme would or would not work.

Marketing and advertisement companies have large experience in predicting consumer behaviour, which can be used in developing both programme and problem theory. Technical literature may also provide insights in mechanisms at work. Programme decision makers (e.g. programme management) and programme workers may be interviewed regarding what they perceive as interesting parameters. Another possibility is to interview the target group of the programme (e.g. using a focus group). The findings can then be analysed and refined by the evaluators.

# 3 EVALUATION PLANNING

This chapter presents a methodology for making initial decisions about programme evaluation strategies. The guidelines make suggestions regarding:

- How to determine the overall **level of effort** and **focus** of the evaluation.
- **Practical issues** to address, to provide the evaluation team with an idea of what level of effort might be required for various evaluation strategies, in light of the amount, type and quality of existing data resources.
- What **limitations** are there on the evaluation due to the timing of the decisions that the evaluation is supposed to support (e.g. how to time the evaluation to allow taking evaluation results into account in programme decision making)?

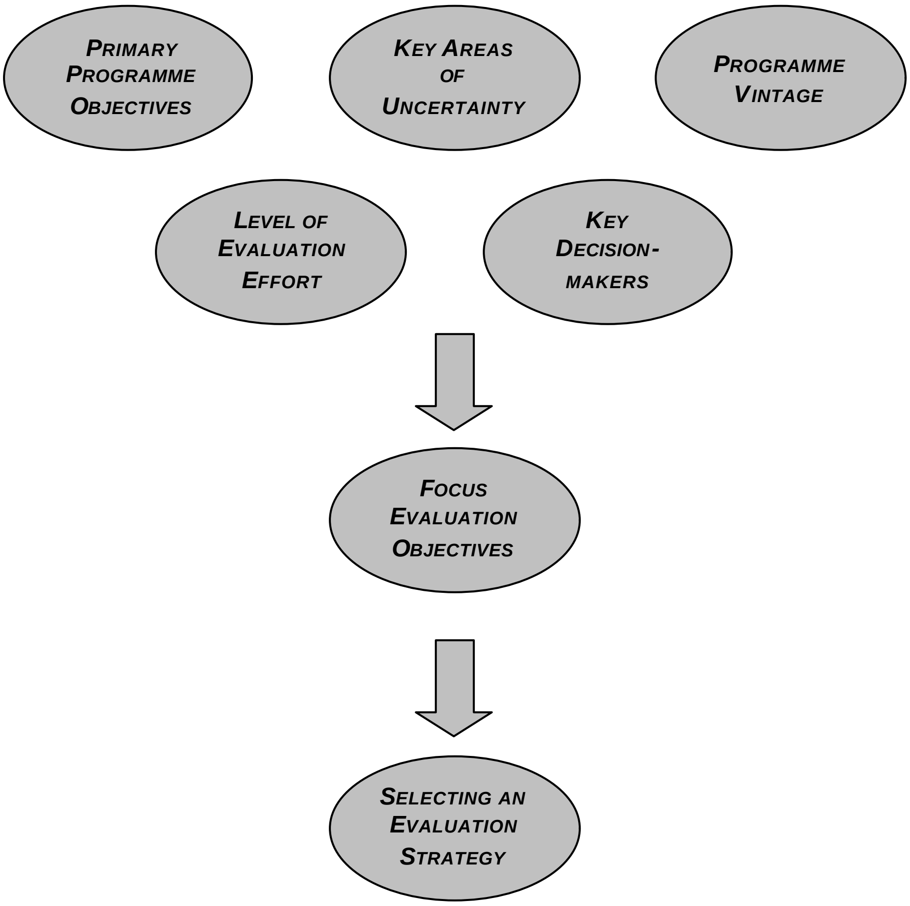
Please refer to Chapter 4 and 6 for specific impact evaluation examples.

Prior to selecting an appropriate evaluation strategy for an EE programme, one should consider the following issues:

- Primary programme objectives – Gain understanding of the programme, its objectives, and how it is supposed to operate.
- Key areas of uncertainty – Identify areas of uncertainty and the impact of making incorrect assumptions about these. Use programme understanding to list key issues to be examined and prioritise the list.
- Programme vintage.
- Appropriate level of effort for the evaluation.
- Key decision-makers regarding the programme and the most important decisions to be made about the programme.
- Focus the evaluation.

Each of these important issues is discussed in more detail below.

*Exhibit 3-1: Issues to consider prior to selecting an evaluation strategy.*




# 3.1 CO-ORDINATION OF PROGRAMME AND EVALUATION

Ideally, the needs of the **evaluation should be incorporated early into programme design**. The advantages of doing this are:

- Evaluation costs can be estimated and included in the programme costs and the benefit/cost reflection of the programme thus providing a more complete picture of the resource consumption.

- Data needs are identified early in the process and thus allow for gathering these at the most convenient and cost-effective time in the programme process. For example, information on participants' background might be relevant in the evaluation and it little or no additional effort to include questions concerning participants' background in a programme questionnaire, which is to be distributed anyway. The cost of preparing an additional questionnaire after programme completion is thus avoided and the participants are not inconvenienced unnecessarily. Another example could be meters or metering, which are included as part of the DSM measure to ensure availability of metered data necessary for the evaluation.

- Areas of uncertainty are clearest at the time of formulating the programme.
- Planning the evaluation at an early stage helps keep programme objectives in clear view during design and implementation.
- Furthermore, it helps focus on how to measure/establish programme success - what are the indicators and what are the success criteria.

Naturally, a need for evaluation of a certain item might first appear after the completion of the programme, However, this is rarely the case if careful planning of both programme and evaluation has been carried out.

## 3.2 EVALUATION PLANNING PROCESS

Having decided to evaluate, the ex-post evaluation planning process should consist of the following actions:

- Decide who at the company is to have the **responsibility** for the evaluation.
- Determine the **purpose** of the evaluation - why do you want an evaluation, what is the goal of the evaluation, and which questions should the evaluation answer.
- What are the **mechanisms** at work in my case? Look at evaluations already done and literature about the relevant mechanisms (e.g. what is the reason people buy non-clean vehicles?)
- Decide whether the evaluation should be **divided** into several smaller studies or performed as one large study. The choice depends on:
    - Whether the evaluation simultaneously touches different areas such as economy, technology, or implementation.
    - Whether the buyer of the evaluation has special interests and opinions about arrangement and presentation of the evaluation of particular areas.
    - Whether the service to be evaluated is divided into several phases/stages.
- **Organise** the evaluation:
    - Decide who is to perform the analysis.
    - Decide who is going to use the results.
    - Contact external entities to be involved in the evaluation.
    - Arrange for good communication between the various parties involved.
    - Make sure that the person responsible for the DSM activity is in complete agreement with everyone involved about the structuring of the evaluation.
- Specify the **general form** of the evaluation (e.g. choice of method/strategy, see Chapter 4 for more detail on this issue).

- Plan the evaluation in **detail** (e.g. select measuring points). The evaluation could for example include profitability, energy efficiency, and how the co-operation worked. Regarding impact evaluations make sure to determine how to establish a baseline and discuss the uncertainty.
- **Implement** the evaluation.
- **Present the evaluation results** so that they are immediately useful to the company and other relevant parties.

Exhibit 3-2 presents an example of a ready-to-copy evaluation plan checklist**{** XE "evaluation plan checklist" **}**, which was prepared by NUTEK**{** XE "NUTEK" **}**, Sweden in 1993.

*Exhibit 3-2: The NUTEK evaluation plan checklist.*

## Evaluation plan for

Project ……………………………………………………………………………………………………………..

**Project manager** …………………………………………………………………………………………..……..

Evaluation plan established by the project manager and …………………………………………………..…….

……………………………………………………………………………………………………………...…….

This form is both a checklist over the establishment of an evaluation plan and a condensation of the plan itself.

**Checklist**: The form is used as a checklist by point after point checking off and giving the date for each decision taken, and jotting down the salient points of each.

**Evaluation plan**: In connection with small evaluations, this form can be the entire evaluation plan or at least the major part of it. With a larger evaluation, this form can be a shorter presentation of the evaluation plan and function as a sort of table of contents.

| Step | | | Date |
|---|---|---|---|
| 1 | The evaluator is ……………………………………………………………………..………….. | | |
| | *Decision taken:* | ❑ | ________ |
| 2 | Goal: We wish to find out the following from the evaluation ..…………………………..…….. | | |
| | ………………………………………………………………………………………………… | | |
| | ………………………………………………………………………………………………… | | |
| | A more complete presentation is to be found in ………..………………………….………… | | |
| | ………………………………………………………………………………………………… | | |
| | *The goal has been discussed, decided, and written down:* | ❑ | ________ |
| 3 | When to evaluate: The evaluation is part of the project plan for the energy service, and the plan is to be found in …………………..…………………………………………. | | |
| | ………………………………………………………………………………………………… | | |
| | *The decision of when the evaluation shall take place has been taken:* | ❑ | ________ |

| **Step** | | | **Date** |
|---|---|---|---|
| 4 | A division of the evaluation into part-studies has taken place<br>Yes ❑ No ❑<br>A complete description of the part-studies may be found in …………………………..<br>……………………………………………………………………………………………<br>*A decision in this matter has been taken:* | ❑ | |
| 5 | The evaluation is organised as follows …………………………………………………..<br>……………………………………………………………………………………………<br>……………………………………………………………………………………………<br>A detailed organisation description is to be found in …………………………………….<br>……………………………………………………………………………………………<br>……………………………………………………………………………………………<br>*The organisation has been decided on and written down:* | ❑ | |
| 6 | The specification of the evaluation in general …………………………………………..<br>……………………………………………………………………………………………<br>……………………………………………………………………………………………<br>……………………………………………………………………………………………<br>A complete description of the general specification can be found in ………………..<br>……………………………………………………………………………………………<br>*The specification has been discussed, decided, and written down:* | ❑ | |
| 7 | a) Detail planning done by ……………………………………………………………….. | | |
| | *The decision in the matter has been made:* | ❑ | |
| | b) The detail plan is to be found in ……………………………………………………… | | |
| | *The decision in the matter has been made:* | ❑ | |
| 8 | *Execution* | ❑ | |
| 9 | The evaluation shall be presented as follows ……………………………………………<br>……………………………………………………………………………………………<br>……………………………………………………………………………………………<br>A more detailed description is to be found in …………………………………………….<br>……………………………………………………………………………………………<br>*The presentation forms have been discussed and decided on:* | ❑ | |

The NUTEK checklist may be expanded to fit your purposes. For example could names of the decision-maker of each step be added to the list. The main point is, that the checklist remains a checklist that provides an overview as supplement to the more detailed evaluation plan.

## 3.3 PRIMARY PROGRAMME OBJECTIVES

The primary programme objective may determine the primary objective of the evaluation effort. There are two basic categories of primary programme objectives:

- **Environmental/energy** resource objectives:
  - Reduce $CO_2$ emissions.
  - Meet future energy capacity needs.
  - Reduce energy imports.
  - Transform markets.
  - Overcome market barriers.
  - Promote general economic development.
  - Develop a strong EE service industry.
- **Business profitability** objectives:
  - Generate profit and increase profit margins.
  - Retain customers.
  - Generate positive public relations.

Typically, the **purpose** of the evaluation is to quantify, to the extent possible, how well the programme is accomplishing its primary objective. This quantification is somewhat easier for emissions, energy use, energy load, and business profitability objectives for which metrics are widely accepted and processes/measurement techniques are available for assessing those metrics.

It is a little more difficult to quantify success in achieving market change objectives, for three reasons:

- Markets involve the complex interaction of numerous forces, making causes of market change very difficult to establish.
- There are often no generally accepted and available metrics which indicate that a specific market has changed.
- Market change tends to happen very slowly, making it more challenging for evaluation to help guide programme implementation and to assess how well programmes are performing. The extended time frame for change also reduces certainty about exactly what is causing observed market changes, as more and more factors have time to influence the market.

As a result of the difficulties quantifying success in achieving market change objectives, evaluators typically collect data on a range of **market indicators** that provide evidence that a market is changing.

The following exhibit illustrates a typical relationship between the programme objective and the primary evaluation objective. To some extent the target group of the programme determines the areas of interest of the evaluation. Therefore these should also be identified.

**Please note, that the evaluation efforts are not necessarily directly linked to the objectives of the programme in question.**

All programmes can benefit to at least some degree from process evaluation activities, in which the efficiency of the programme and the nature of participating and non-participating consumers/businesses can be analysed, so that programme costs can be minimised and the amount of impact per unit of expenditure can be maximised.

Other examples of evaluation objectives not directly related to primary programme objectives are:

- Cross comparison of different programmes to establish which programmes are likely to continue in a competitive market.
- Understand or develop the programme tracking/monitoring system, so that participation can be documented (see below for more detail) and compile and analyse the usefulness of existing data resources of value to the programme evaluation.
- Determine whether opportunities exist for programme implementers to collect data of value to the evaluation less expensively and more efficiently than for evaluation researchers to do so. Determine the practicality of such data collection and its impact on both other activities implementers must perform and the validity of the evaluation results.
- Assess the energy savings engineering algorithms and the assumptions and supporting data behind them (see next section).

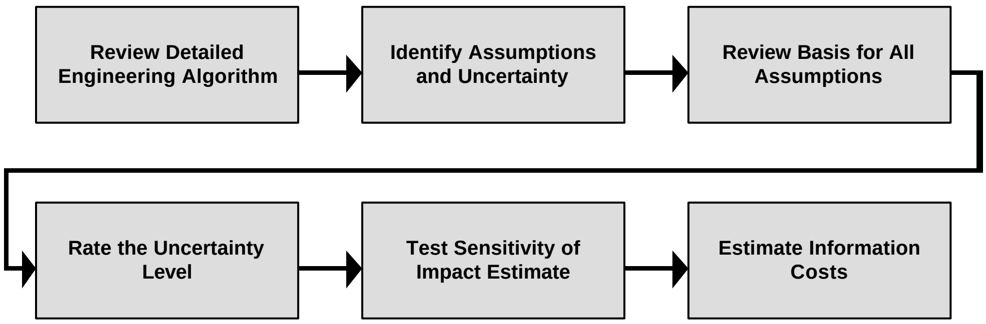
*Exhibit 3-3: Relationship betweeen programme objectives and primary evaluation objective.*

| **Primary Programme Objective** | **Primary Evaluation Objective** |
|---|---|
| Reduce $CO_2$ emissions | Change in $CO_2$ emissions/change in amount of energy used |
| Meet future energy capacity needs and/or reduce energy imports | Change in energy use:<br>- Amount of energy use (e.g., kWh)<br>- Capacity of energy use (e.g., kW) |
| Transform markets/overcome market barriers | Change in market indicators — percentage of retailers stocking the energy efficient product, percentages of consumers and businesses aware of the product, product penetration, percentage of businesses engaging in the efficient practice, etc. |
| Promote general economic improvement | Change in market indicators — number of new business starts, level of retail sales for discretionary products, housing starts, etc. |
| Develop a strong EE services industry | Change in market indicators — number of ESCOs, annual revenues for ESCOs, percentage of businesses aware of ESCOs, percentage of businesses likely to contact an ESCO in the next year, etc. |
| Generate profits for sponsors/investors | Change in sales and profit margins |
| Retain customers | Change in retention rate, change in per-customer margin |
| Generate positive public relations | Change in market indicators — Awareness of firm among targeted customers, number of mentions in various media, percentage of positive mentions in various media, etc. |

# 3.4 KEY AREAS OF UNCERTAINTY

One can always learn more about a programme through evaluation activities. However, the design and performance of an evaluation comes at a cost. To maximise the usefulness of evaluation results and minimise evaluation cost, decision-makers must determine how much uncertainty**{** XE "uncertainty" **}** they are willing to accept and when to require additional analysis and data collection to refine and confirm estimates. This includes asking the question: What are the consequences and the alternatives?

*Exhibit 3-4: Course of action for review of programme plans.*




We suggest the following course of action, when reviewing programme plans to develop an impact evaluation strategy:

- Review or develop a detailed engineering algorithm or set of algorithms for estimating programme impacts. Ideally, these algorithms are already included as part of the programme's tracking system, so that the tracking system can generate an estimate of the programme's (or each participant's) impacts. The algorithm should include all factors that may influence savings, such as free-ridership, spill-over, percentage of the load/capacity savings that is coincident with the utility's system peak (for utility load-reduction programmes), seasonal differences in hours of use, persistence, etc.

- Clearly identify all assumptions and areas of uncertainty in all programme plans.

- Review documentation for (or document) the basis for all assumptions. If a specific assumption is based on the results of previous evaluations, indicate the degree to which key programme conditions differ or are the same as the current programme (type and composition of participant population, year of data collection and analysis, economic environment, industry structure, etc.).

- Rate the uncertainty level around each term of the algorithm as high, medium or low, and document the logic of this rating.

- Test the sensitivity of the impact estimate - and its effect on programme cost-effectiveness - to changes in various terms of the engineering algorithm. It may not matter that there is great uncertainty in certain terms of the engineering equation, if wide variances in these terms have only an insignificant effect on the programme impact estimate and cost-effectiveness.

- Estimate the cost of obtaining information that would *significantly* reduce the uncertainty of the algorithm terms having high or medium uncertainty and for which the sensitivity analysis showed that variations produce a significant change in the impact estimate. It may be prohibitively expensive to significantly reduce the uncertainty around certain terms, for example, if it means a very large number of EE installations must be fully metered. However, it may be relatively inexpensive to address free-ridership issues when surveying programme participants.[3]

Having completed these steps, those responsible for the evaluation methodology have the information needed to make an **informed judgement** with regard to matching evaluation resources to the areas of greatest uncertainty and greatest possible effect on the evaluation's impact estimate.

**Case Example: Evaluation of the Energy Efficiency Check**

We realised that the goal of our evaluation was not compatible with the available information and resources. The initiating question of the evaluation was thus changed from "How much energy has been saved?" to "How many energy saving measures have been implemented due to the energy check?".

**Norsk Enøk og Energi AS, Norway**

**Process evaluation** may encompass one or more elements of the programme process from planning and design to delivery. It constitutes a type of administrative evaluation and final conclusion on the tracking and monitoring efforts carried out during the programme period.

Sometimes the process evaluation is split in two: An external and an internal part. The external process evaluation focuses on the part of the process involving external parties while the internal process evaluation focuses on the internal procedures. External process evaluation includes trade ally research (interviews, surveys, focus groups**{** XE "focus groups" **}**) and customer research (both participating and non-participating). The internal assessment includes review of programme database and tracking system, review of programme guidelines, status reports, and marketing material, as well as interviews / focus groups with programme staff.

Some of the uncertainties, which can be addressed by process evaluation, are:

- Objectives - Did the programme meet them? Have they changed? Why/Why not?

- Communications at all interface points (with trade intermediaries, customers, between utility and corporate staff) - Was it accurate, timely, and sufficient?

- Programme design process - Was it successful? How can it be improved?

- Programme data tracking (content and structure) - Quality? Is it accurate? Can it be streamlined? Does it support evaluation?

---

[3] *A detailed description of a rigorous, quantitative way to use a similar evaluation planning approach appears in a 1993 paper on allocating evaluation resources: "A Framework for Strategic Evaluation Planning," by Kurt Kiefer, Wisconsin Power & Light Company, Proceedings of the 1993 International Energy Programme Evaluation Conference, Chicago, IL, USA, August 1993, CONF-930842, pp. 623-628. Other papers in the same volume also address this and similar topics, pp. 629-648.*

- Customer satisfaction - How satisfied are customers? How can satisfaction be increased?

- Programme delivery - Are there any bottlenecks? Are there adequate resources at each stage? How does actual delivery compare to planned delivery?

See also Section 2.4.2 on questions, which the evaluation may help answer.

# 3.5 PROGRAMME VINTAGE

Evaluation needs change as programmes mature. New programmes tend to need more exploratory research, while more mature ones require only very targeted studies about unresolved issues.

As a rule of thumb, an overview of the main evaluation needs by programme vintage**{** XE "programme vintage" **}** is presented in Exhibit 3-5.

*Exhibit 3-5: Overview of evaluation needs by programme vintage.*

| **First Year Programmes{** XE "first year programme" **}** |
|---|
| Review Tracking & Monitoring System |
| Conduct Interviews |
| Document Actual Operation |
| Determine Site to Meter and Install Meters |
| **Second & Third Year{** XE "second & third year programme" **} and Pilot Programmes{** XE "pilot programme" **}** |
| (Review Tracking & Monitoring System) |
| (Conduct Interviews) |
| (Document Actual Operation) |
| (Determine Site to Meter and Install Meters) |
| Comprehensive Review of Tracking System |
| Comprehensive Impact Evaluation |
| Process Evaluation |
| **Mature Programmes** |
| Brief Interviews |
| Review & Update Programme Documents |
| Comprehensive Review of Tracking System |
| Limited Research to Remove Uncertainties |
| Estimate Programme Impact and Cost-effectiveness |
| Review Evaluation Results with Planners |

### 3.5.1 FIRST YEAR OF PROGRAMME IMPLEMENTATION

- Review the programme tracking/monitoring system:
  - Ensure that it is accurately capturing key information needed to make implementation management decisions about the programme and to later evaluate the programme's impacts.
  - Conduct a cursory analysis of who is and is not participating in the programme (if programme managers have not already done this) and alert programme managers/implementation staff to unexpected findings.
  - Examine key performance indicators to identify participation or programme impact issues. Decide whether limited survey or focus group research is warranted to provide a basis for making mid-year refinements to the programme.
- Conduct interviews with programme planners, managers, and implementation staff:
  - Determine how the programme actually is being implemented (as opposed to how it was planned to operate).
  - Identify bottlenecks in the programme delivery process, procedural issues, and communications problems.
  - Review results of tracking system review.
- Document how the programme actually operates. Include a discussion of the programme process, including:
  - Development and design of the programme (the development process as well as the programme structure and impact engineering algorithms).
  - Marketing efforts.
  - Consumer/business contacts.
  - Participation processing.
  - Participant data collection.
  - EE measure implementation.
  - Other key activities/points of interaction (such as interactions with implementation contractors or other trade allies).
- Determine sites to meter, and install metering equipment (if appropriate and not already in place).

### 3.5.2 SECOND/THIRD YEAR OF PROGRAMME IMPLEMENTATION AND PILOT PROGRAMMES

- Perform any tasks listed above for **first programme year** that were not performed during first year.
- Conduct comprehensive review of the programme tracking/monitoring system:
  - Review accuracy of and trends in data.
  - Examine key performance indicators to identify participation or programme impact issues to be addressed in survey or focus group research.
  - Perform market evaluation analysis — Identify and profile participating and non-participating populations, segments of high and low impact, etc.
  - Design final sample design plan for participant and non-participant survey/metering research as appropriate. Select samples as appropriate.
- Conduct comprehensive **impact evaluation**:
  - Implement primary impact evaluation strategies, including survey research, secondary research if appropriate, and metering data collection, as well as any additional meter installations needed.
  - Analyse all data collected, to estimate programme impacts and values for specific impact parameters.
  - Conduct programme benefit/cost analysis.
  - Explore reasons for specific impact evaluation findings/results. Compare results to the programme's engineering algorithms, including assumptions behind specific algorithm parameters. Work with programme planners to resolve inconsistencies and anomalies.
- Conduct process evaluation:
  - Implement process evaluation strategies, including interviews with programme staff (unless performed recently), secondary research (if appropriate), and research with consumers targeted by the programme, both participating and non-participating.
  - Analyse data collected.
  - Report results to programme management in constructive manner.

### 3.5.3 MATURE PROGRAMMES

- Conduct brief interviews with programme staff, to identify programme changes.
- Review and update programme documentation.
- Conduct comprehensive review of the programme tracking/monitoring system:
  - Review accuracy of and trends in data.

  - Examine key performance indicators to identify participation or programme impact issues to be addressed in survey or focus group research.

  - Design final sample design plan for participant and non-participant survey/metering research as appropriate. Select samples as appropriate.

  - Perform market evaluation analysis — Identify participating and non-participating populations, segments of high and low impact, etc.

- Conduct limited, targeted research with targeted population, to resolve lingering uncertainties and verify implementation, as needed. Possible areas include:

  - Free-ridership, spill-over and persistence (which are apt to change over time).

  - Issues identified in performance indicator review and programme staff interviews.

  - Remaining market potential and ways to increase penetration (e.g., through programme redesign).

  - Selected impact parameters.

- Estimate programme impacts and cost-effectiveness.

- Review selected evaluation results with programme planners (as appropriate).

As the method described above indicates, one need not conduct a comprehensive evaluation of a programme every year it is being implemented. Instead, evaluation should be used as a tool for addressing areas of uncertainty at each point in a programme's life. Still, it is important to (1) conduct at least one comprehensive evaluation of every major programme and especially (2) conduct research in the market place, to know what is really happening in the programme.

## 3.6 DETERMINE THE GENERAL LEVEL OF EFFORT FOR THE EVALUATION

As discussed above, it is usually not feasible or cost-effective for every programme to receive the same level of analysis each year. At the same time, it is not possible to indicate which share the evaluation cost should make of the total programme cost. However, the following guidelines are offered as a rationale for determining the general level of effort for an evaluation of a specific programme:

More **comprehensive evaluations** are generally worthwhile to perform if a programme …

*. . . is expected to yield very significant impacts and/or has cost a great deal to implement.* It is only logical that, when significant resources are at stake, careful attention should be given to how a programme is operating and what its impact is.

*. . . is a pilot programme.* Pilot programmes serve as the trial stage for larger, more extensive, more costly programmes. Comprehensive evaluations of such programmes may result in significant cost savings or impact increases when full-scale programmes are later implemented.

*. . . involves a new programme delivery approach or a new technology.* For programmes that are significantly different from programmes implemented in the past, it is likely that important

lessons will need to be learned and that certain preconceptions of programme planners and implementers will be inaccurate. This is particularly true for entirely new types of programmes (e.g., market transformation programmes) or new technologies with which organisations may have no prior experience.

. . . *will receive a high degree of scrutiny.* This is self evident. The level of scrutiny is likely to increase if (1) the programme or its sponsor/implementing organisation are viewed with scepticism by entities having some political clout, (2) significant sums of money are at stake, or (3) significant public relations capital is at stake (possibly the case with regard to programmes targeted at reducing greenhouse gas emissions so that governments can meet international commitments).

Simpler, **less expensive evaluations** should be considered for less controversial, lower-impact, lower-cost, and/or more commonplace programmes.

Of course, the level of effort for the evaluation may be strongly determined by the overall level of funding available for the programme's design, implementation and evaluation. Those charged with developing and implementing EE programmes generally want as much of programme funding to go toward achieving programme impacts as possible, rather than precise measurement of what these impacts are. The uncertainty analysis described above may help decision-makers appreciate the value of addressing specific evaluation issues and the possible cost consequences of not doing so.

*Exhibit 3-6: Rules of thumb regarding level of effort for the evaluation.*

| **Comprehensive Evaluation** | **Less Comprehensive Evaluation** |
|---|---|
| if | if |
| Significant Funding | Moderate/limited Funding |
| Significant Expected Impacts | Moderate/limited Expected Impacts |
| Pilot Programme | Well-known Delivery or Technology |
| New Delivery or Technology | Not Controversial |
| Subject to Scrutiny | |

## 3.7 KEY DECISION-MAKERS

The{ XE "key decision-makers" } individuals and organisations needing to make decisions about the programme help determine both the relative importance of various aspects of the evaluation and the level of rigor with which the evaluation must be performed. The evaluator must determine:

- Who must make decisions about the programme?
- What aspects of the programme will affect those decisions?

Remember also to establish **when** decision-makers must make their decisions. Important, comprehensive evaluation data may be of little use, if they come to decision-makers after key decisions have been made.

### 3.7.1 MANAGERS OF PUBLIC POLICY PROGRAMMES

Typically, the closer the decision-maker is to the individuals charged with programme implementation, the more **discretion** the decision-maker has regarding evaluation requirements and the more variability there is in the focus of evaluation activities.

Consider a programme manager at a utility, which is implementing an EE programme. The manager may have a great deal of experience in business or industry, or with regard to a specific technology, leading to a strong degree of comfort about the programme's assumptions for certain impact parameters such as the number of hours that certain equipment is used daily, or the type of equipment likely to be replaced by more efficient equipment. However, the manager may have some uncertainty about the efficiency of equipment that would be purchased in the absence of the programme or the effect of weather on equipment energy use. The manager would want the evaluation to focus minimal resources, if any, on the "understood" impact parameters and more resources on those about which there is greatest uncertainty. That is, the manager's decision-making needs would e.g. be met by an evaluation focused on the effect of weather on equipment use.

However, the manager's opinions may not be shared by others, especially those with less or a different type of experience in the market. Perhaps the manager had a strong experience that left an indelible impression several years in the past, and perhaps the industry or equipment has changed somewhat in the intervening years. The manager's perspective may be idiosyncratic, if not inaccurate.

If the **impetus for the programme** has come from within the organisation, this perspective may not be challenged for some time. The manager must merely convince upper management that the programme is achieving its objectives and, failing strong evidence to the contrary, upper management may not ask detailed enough questions or have sufficient technical/industry knowledge to become concerned that the manager's claims are not rigorously supported. After all, upper management may only be concerned with high-level questions such as "Should this programme be continued?" or "Should this programme be changed?"

### 3.7.2 OTHER DECISION-MAKERS ASSOCIATED WITH PUBLIC POLICY PROGRAMMES

On the other hand, decision-makers other than the programme manager may affect the nature of the evaluation:

- Perhaps the organisation is faced with a need to dramatically reduce costs and all possible cost-cutting opportunities are examined. Basic assumptions about the wisdom of implementing the programme may be challenged. Or, if the programme costs are large enough, other decision-makers may want to examine ways to improve programme cost-effectiveness, i.e., achieve the same or similar results by spending less.

- Perhaps some in upper management require strong documentation as a matter of policy or style. The evaluation may be required to produce data that confirm or refute the programme manager's assumptions.

- Perhaps the programme's impetus comes from an overseeing external source – government, regulators, etc. – which is either paying the programme manager's organisation to implement the programme or requiring it to do so. Decision-makers in these organisations may be responsible to a wide range of constituency, some of which may be hostile to the programme manager's organisation. Perhaps the decision-makers, themselves, are hostile to the organisation. In this scenario, the evaluation must be thorough enough and the evaluation results rigorous enough to withstand scrutiny by possibly hostile reviewers.

- Perhaps a government agency required the programme (or programmes like it) to be implemented with the intent of showing the rest of the international community that it is meeting its commitments to reduce greenhouse gas emissions. Now, decision-makers in other countries must be convinced of the validity of programme impact claims made by the implementing organisation. Typically, the requirement is for the evaluation to meet some sort of universal standard, most likely one that is both universally accepted in the academic community but practical enough (in terms of cost and data requirements) for all countries involved to meet.

Regardless of who the decision-makers are for public policy EE programmes, the evaluation team should confirm, from the beginning, what the primary programme objectives are with the decision-makers. It will also be helpful to obtain the decision-makers' initial assessment, if any, of the indicators for which specific changes would constitute the strongest evidence of programme-induced market change.

### 3.7.3 DECISION-MAKERS IN ENERGY SERVICES PROJECTS

Consider a similar process in the case of an energy services company proposing to implement an EE measure of some sort (lighting change-out, HVAC system upgrade, etc.) for a potential client. Decisions about how to estimate the energy impacts of specific EE measures may become much more simplified.

In this case there are two main decision-makers. The ESCO's management seeks a specific profit margin on the project and will price the project and any related evaluation (i.e., monitoring and verification) accordingly. Evaluation becomes a tool for increasing client comfort with the project. The client's management is likely to want to spend as little as possible to reap the financial rewards of the EE project, and knows that the evaluation (i.e., monitoring and verification) represents a cost that must be subtracted from the financial benefits it will receive.

Evaluation then becomes a method for **addressing risk**. What level of risk does the client feel is inherent in the project? Are engineering assumptions behind the pre-implementation estimate of the project's savings straightforward and generally accepted? How sophisticated are the client's technical staff? How comfortable are they with the impact estimate? Are they willing to take responsibility for savings estimate as long as the ESCO is held accountable for a specific level of implementation quality?

Perhaps the client understands the technology well enough so that no ongoing monitoring is necessary, but rather only verification that the measure is being installed competently. Or perhaps the client perceives some risk and, rather than requiring sophisticated monitoring and evaluation, prefers to off-load risk through a shared savings agreement based on some bottom-line indicators the client is confident will reflect the project's success. In this way both the ESCO and the client have a strong interest in the project yielding maximum savings.

Yet another scenario is one in which the client is a government body negotiating work to be done on a government-owned facility. This entity may be required to have monitoring and verification for the project follow widely accepted guidelines such as those appearing in the **International Performance Measurement & Verification Protocol** for energy services projects (see bibliographic summary in Appendix B). Whatever the case, in the ESCO project scenario, evaluation becomes another part of a contract between two parties, and whatever the two parties are willing to accept and agree to determines what is required.

Energy marketers implementing EE programmes as a strategy for generating profit, or for retaining existing customers/attracting new ones have a similar decision-making focus to that of ESCOs, because, in effect, they are functioning as ESCOs. Again, what is most important is customer satisfaction, and decisions about the rigor and cost of evaluation activities (monitoring and verification) are driven by what is required to maximise satisfaction and profit. The value of an evaluation to the energy marketer is in providing data related to customer retention, profit margins, profitability, market share, market position, and other competitive issues, rather than energy savings per se. Energy savings are only a mechanism for addressing these more important issues. As such, evaluation of energy impacts of specific projects (or of all projects within a specific time period) is secondary and a contractual issue primarily serving the needs of the marketer's client's decision-makers.

### 3.7.4 DECISION-MAKERS IN UTILITIES OPERATING LOAD MANAGEMENT PROGRAMMES

The key decision-makers with regard to programmes designed to delay or avoid the need for costly upgrades to T&D equipment include the programme manager, system planners, and, if the utility is subject to regulation, the regulating authority. For this type of programme, more than any other, decision-makers are likely to require compelling, easy-to-understand evidence of programme impacts, because the reliability of portions of the T&D system is at stake.

The central issue is whether the programme has **delayed the need for the planned upgrades** and, if so, for how long. Answering this question implies a careful analysis of the assumptions and analysis used by those planning the upgrades. As noted earlier, the very process of closely analysing these assumptions can sometimes result in significant cost savings and delayed implementation of upgrades. However, the evaluation will need to assemble strong evidence, with conservative assumptions, to support claims of any real energy savings. Data from measurements at the targeted transformer or substation level are likely to be considered the strongest source of impact evidence, and system planners and the evaluation team will need to work together to provide a reasoned estimate of the reliable impact of the programme.

## 3.8 FOCUS THE EVALUATION

Given all the possibilities and buts and ifs, it is easy to get side-tracked in the evaluation efforts. It is therefore important to focus**{** XE "focus" **}** before proceeding any further with the evaluation.

A simple method to breaking down the evaluation into manageable pieces is to go through the five steps below:

1) What is the **question**, which the evaluation seeks to answer (initiating evaluation question)?

2) What does the question **aim** to clarify?

3) What are the underlying **assumptions** (with respect to mechanisms)?

4) Which **indicators** will be used to answer the question?

5) What is the **approach** for establishing the indicator?

If it becomes difficult to answer the questions in step 2-5, then the first question is not specific enough. It may be necessary to use iterative thinking to arrive at a sufficiently specific initiating question. Step 3 relates the programme theory and the problem theory discussed above.

An example may illustrate the idea more clearly. We therefore consider a programme example. The programme objectives were to provide training in heating systems and EE improvement of such to energy managers in all industry. The training course and certificate were free of charge and only offered to employed energy managers. Direct mail was used to promote the course. A regional energy centre offered the programme and the training was outsourced to a specialised consulting company. The course was repeated four times in a row and lasted a week each time.

All elements of the programme may be subjected to evaluation and they are not all equally easy to evaluate. Some examples of **initiating evaluation questions** (Step 1) are presented below (by far not a complete list):

- Would others be interested in the course (and to a greater energy impact)?
- Was direct mail a good way to create interest?
- What energy savings were achieved due to the programme?
- What was the motivation of the energy managers to participate (were they planning to improve their systems anyway)?
- What is according to the energy managers the driving force in initiating improvement of the heating system efficiency?
- Which industrial branches did the participants represent?
- What impact does the promised certificate have on participation rate?
- Would it be more cost-effective to use regional energy centre staff as trainers?

Some of the questions may need to be divided into sub-questions to allow sound, structured evaluation (e.g. what is *good* training?). Careful choice of words is necessary to direct the evaluation in the wanted direction. Exhibit 3-7 shows an example, which illustrates that the same question may be asked for different reasons. The example should not be seen as “an absolute truth” – it only aims to hint at what might be of interest and to point out that focus and explicit, precise formulations are important.

*Exhibit 3-7: Example of breakdown of an evaluation.*

| **Question:** | Did the consulting company fulfil its contract to our satisfaction? | | |
|---|---|---|---|
| | **Interpretation 1** | **Interpretation 2** | **Interpretation 3** |
| **Aim:** | Form a basis for a decision on whether or not to employ this consulting firm again in this or other training courses. | Release payment for services rendered. | Did the contents and structure of the training course give new and useful skills to participants? (The exact skills wanted could be listed.) |
| **Assumptions:** | The consulting firm will perform in the same manner in similar and different training courses (including applied approach and choice of teacher). | Contract fulfilment can be measured in number of hours, subjects taught, and the staff's impression of the course. | ▪ The participants have not been exposed to other ways of obtaining these skills during the course period and two months forward.<br>▪ If the skills are useful, companies will send additional people to the course.<br>▪ The participants will apply their skills no later than 2 months after the programme. |
| **Indicator:** | ▪ List of subjects covered in the training.<br>▪ Participant small talk during breaks.<br>▪ The level of satisfaction of the participants including their suggestions for improvement at the end of the course and later.<br>▪ Number of energy managers, who have applied their newly acquired skills.<br>▪ Consultant's impression of the course. | ▪ List of subjects covered in the training.<br>▪ Staff's impression of the course.<br>▪ Invoice from consulting company. | ▪ List of subjects covered in the training.<br>▪ Number of energy managers, who have applied their newly acquired skills.<br>▪ Skills, which have been applied.<br>▪ Number of companies, which have sent other energy managers to the course. |
| **Approach:** | ▪ Review subjects listed in course material.<br>▪ Interview staff that was present during training.<br>▪ Participant comments overheard by staff or expressed to staff.<br>▪ Review assessment forms completed by participants at the end of the course.<br>▪ Interview x% of the participants 2 months after completion of the course.<br>▪ Interview the teachers. | ▪ Review subjects listed in course material.<br>▪ Interview staff that was present during training.<br>▪ Compare invoice with teacher presence during course and the amount of course material distributed. | ▪ Review subjects listed in course material.<br>▪ Interview x% of the participants 2 months after completion of the course.<br>▪ Compare course lists to find "repeat" companies. |

**Case Example: Evaluation of a CFV Programme – Powershift**

The evaluation of the PowerShift programme currently employs a number of performance indicators ranging from purely quantitative carbon and regulated atmospheric emission savings and cost-effectiveness indices to more qualitative and non-emission based indicators such as awareness, cost differentials and the development of refuelling infrastructure.

The Energy Saving Trust is required to report to Government annually on the total savings and policy cost per tonne of $CO_2$, total green house gas, $CH_4$, $NO_x$, and particulate matter. Both savings and policy cost are calculated on an annual and a lifetime basis. The policy cost equates to government funding and is effectively the total programme expenditure less partner and customer contributions. Emission factors used are sourced from the report of the alternative fuels group of the Cleaner Vehicles Task Force.

Other **prime indicators** monitored are:

- CFV sales per year;
- Total vehicle populations;
- Number of refuelling stations;
- Financial price premiums between CFVs and conventional equivalents;
- Residual values of vehicles.

The **secondary indicators** include, for example:

- Number of grant applications;
- Number of workshop delegates;
- Number of press articles;
- Number of hotline calls;
- Number of website visitors;
- Number of approved vehicle manufacturers;
- Number of approved converters;
- Number of fuel suppliers.

Indicators are monitored using market research involving vehicle manufacturers, converters, and fuel suppliers.

**Energy Saving Trust, United Kingdom**

## 3.9 REQUIRED BUDGET FOR EVALUATION

It is important to allocate sufficient budget{ XE "budget" }, staff and time to evaluate competently important programmes and projects. In general, evaluations planned early in the programme's life cycle yield more accurate and useful results at a lower cost than those planned and implemented after programme completion.

The appropriate size of the budget for evaluation varies significantly depending on the programme type, the programme objectives, and the evaluation objectives. Roughly estimated, a sound evaluation budget constitutes around 3-10% of the total programme costs.

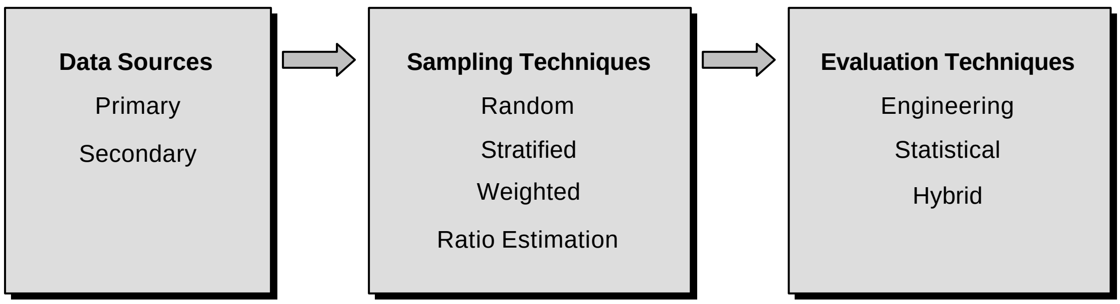
Some flexibility should be designed into the budget, particularly when planning new evaluations, as it may turn out certain components of the evaluation study may go over or under budget. Another point worth remembering is that it might be possible to combine the evaluation efforts of several programmes and/or projects and thus reduce both costs and effort.

# 4 OVERALL IMPACT EVALUATION STRATEGIES

This chapter provides:

- A listing of **resources** required for obtaining data needed in impact evaluation;
- Descriptions of **specific evaluation strategies** that can be used at various levels of effort, from simple and often quite imprecise methods, to highly complex, expensive and much more accurate methods;
- A discussion of the **strengths and weaknesses** of main evaluation strategies, including the types of programmes most appropriate to evaluate with them;
- A discussion of **programme cost-effectiveness**.

*Exhibit 4-1: Impact evaluation strategies.*




## 4.1 IMPACT EVALUATION DATA SOURCES

### 4.1.1 TRACKING AND MONITORING SYSTEMS

Tracking**{** XE "tracking" **}** and monitoring**{** XE "monitoring" **}** systems provide information for impact evaluations as well as process evaluations.

Tracking and monitoring are both a sort of "running" evaluations. Tracking refers to the recording of programme data while monitoring refers to the general surveillance of the programme management, which may lead to adjustments. Often no distinction is made between the two terms.

Tracking and monitoring systems collect programme data on a regular basis providing an ongoing reading of the programme performance useful for programme management and stakeholders. Tracking and monitoring data can thus be used to identify problem areas as well as areas in need for more thorough evaluation. Specific performance thresholds are established before commencing the programme implementation. Also specific evaluation steps to be carried out if threshold values are exceeded are identified prior to programme implementation.

The data collected through the tracking and monitoring systems is also very valuable to the ex-post evaluation of the programme. If the development of tracking and monitoring systems is co-ordinated with the evaluation planning, the total effort needed for data collection may be minimised.

> **Case Example: Energy Efficiency Standards of Performance**
>
> Energy Efficiency Standards of Performance (EESoP) were introduced in 1994 in England and Wales and a year later in Scotland. The promoted EE measures include low energy lamps, domestic appliances, insulation, heating systems, and combined power and heat installations. An essential part of EESoP scheme has been to monitor the EE programmes in three fundamental areas: Energy, customer satisfaction, and quality.
>
> In order to compare the actual energy savings related to heating and insulation measures with predicted savings, a methodology was devised in which the first step was to estimate how much energy could be saved from various measures in each property type, assuming standard heating patterns.
>
> For each project, a sample of at least 5% of properties was selected for monitoring. Meter reading data for a year before and after installation of insulation measures were analysed, taking care to eliminate estimated readings or other anomalous data. Readings were adjusted to take account of weather variations nationally and from year to year.
>
> Results show that individual properties may save much more or less energy than the average predicted by the computer model. The reasons for this depend upon a wide range of factors including occupancy, heating patterns, ownership of electrical appliances and construction details. However, although individual differences may be large, on average they are not significant enough to demand modifications to the existing model. Furthermore, the variations do not relate to customer perception or acceptance of the programme and do therefore not justify a modification of the programme (e.g. alternative marketing approach or different target group).
>
> Using a questionnaire, customer satisfaction with promoted EE measures is monitored by energy suppliers. A minimum of 5% of homes for all measures installed are monitored except for compact fluorescent lighting where 1% or 1,000 customers are monitored (whichever is the less).
>
> Quality of installation is surveyed and reported on by energy suppliers in conjunction with local authorities in a minimum of 5% of homes receiving fixed “fabric” measures such as insulation or heating measures. This quality monitoring checks whether or not the measures have been installed in line with approved British/European Standards etc. For CFL schemes, quality criteria are fulfilled if lamps included on an approved list are used. For appliance schemes, assuming that all products used have relevant CE marking, there are no additional quality monitoring requirements.
>
> So far, customer satisfaction and quality have been satisfactory.
>
> **Electricity Association, United Kingdom**

Examples of possible functions of tracking and monitoring systems are:

- Comparing at regular intervals recorded progress against goals (kW and kWh impacts, participation, or penetration of measures);
- Comparing at regular intervals recorded expenditures against budgets;
- Monitoring backlogs;
- Monitoring contractor activities;
- Supporting reporting requirements of upper management and regulatory staff;
- Tracking effectiveness of promotional approaches;

- Characterising participants;
- Supporting ongoing benefit/cost analyses.

When using tracking and monitoring results, the evaluator should differentiate between findings uncovered by analysis of database and "anecdotes", to ensure that factual information does not get confused with judgmental information. All the same, a comparison of the two will provide valuable information on the factual and the perceived programme performance. Discrepancies could for example identify a need for better communication.

### 4.1.2 PRIMARY DATA SOURCES

The primary sources of data{ XE "primary data sources" } for impact evaluations are presented below. The more complete and more accurate each of these data source are, the less expensive and more accurate the impact evaluation can be.

**Programme tracking system data** – Data about each participant in the programme, which can include contact/address information, account number, details about the measure(s) installed and equipment replaced (if any), demographic/business type information, facility characteristics, appliance holdings, and preliminary estimates of the participant's energy impacts. Tracking system data quality is extremely important; it is the main record of what is happening in the programme. Also of high importance, for statistical billing analyses, is the ability to link a particular participant with that participant's energy usage (e.g., via the account number).

**Surveys with programme participants and/or non-participants** – These can be performed via mail, telephone or in-person, and require market research expertise so that survey responses actually reflect the exact information the interviewer wanted to determine. They are used in most evaluations.

**On-site facility surveys** – These may be considered a subset of surveys with programme participants and non-participants, but they focus on the consumer's facility, home, appliances, and/or energy-using equipment. They are also specifically used to verify that EE measures have been implemented. Usually, they are combined with surveys of the consumers, which can often be performed at little to no additional cost (the cost of placing the researcher at the facility has already been borne) and many surveys may take a relatively short time to complete, compared to the time it takes to complete a facility audit or efficiency measure verification.

**Energy use/billing data** – These data are collected using meters located at the consumer's facility and serve as basis for energy bills to be sent to utility customers. They are one of the central data resources used in billing analyses (and, sometimes, the only one). Higher quality (cleaner) data will speed up the evaluation process and help ensure that precision of the impact estimate is maximised.

**End-use metered data** – These data are collected from instruments that measure energy usage by a specific piece of equipment or process, at the consumer's site. The measurement instruments can range from relatively inexpensive, easy-to-install "loggers" that record only the number of hours a piece of equipment is "on," to expensive, multi-channel devices that take significant skill to install and can record a number of different types of data (temperature, energy use, load level, on and off times, humidity, etc.) for more than one piece of equipment. Key to the use of end-use metering equipment is ongoing verification that the meters are calibrated and collecting data accurately. Even simple light loggers involve the expense of an installation visit; that expense is wasted if the unit malfunctions or for some reason the data it collects are not usable.

End-use metering is most useful for:

- Evaluation of gross savings.
- Estimation of coincident peak load savings.
- Focused evaluation of large-impact facilities that are not well suited to statistical analysis.
- Providing data for statistical analyses and combination approaches.
- Evaluations of specific technologies (gross consumption change for specific measure).
- Addressing specific research issues, such as determining operating hours or estimating interactive effects.
- Joint utility projects where the relatively high costs can be spread across utilities.

### 4.1.3 SECONDARY DATA SOURCES

Secondary data sources**{** XE "secondary data sources" **}** per definition already exist and are available from government offices, previous research studies, equipment manufacturers, research organisations, private firms, etc. From the perspective of the evaluation, they fall under the category of "the best information available" and are used:

- To help confirm the reasonableness of impact estimates or engineering assumptions.
- To provide estimates for needed data items that are otherwise impossible or too expensive to collect (equipment operating efficiencies under different load conditions, equipment sales for specific geographic areas, etc.).
- To help explain unexpected impact findings or supply information on such impact parameters as free-ridership, spill-over, rebound, or persistence.

## 4.2 EVALUATION TECHNIQUES

Evaluators have two basic types of evaluation techniques at their disposal: **Engineering methods{** XE "engineering methods" **}** and **statistical methods{** XE "statistical methods" **}**. Engineering methods are based on defining the basic physical relationships that exist between the change in energy use and the factors that determine that change, based on engineering principles. Statistical methods use recorded consumption data, and compare changes in the level of energy consumption for two populations (e.g., programme participants and a control/comparison group) to isolate the energy impacts of the adoption of the specific EE measures promoted by a programme.

The two techniques can also be combined, the so-called **hybrid technique**. Statistical methods can be enhanced by including an initial engineering estimate of the energy savings for each participant as an explanatory variable in the regression equations**{** XE "regression equations" **}**. The statistical model then estimates coefficients that represent not energy savings but realisation rates**{** XE "realisation rates" **}** (the factor by which the engineering estimates must be adjusted in order to reflect the true energy impact). The usefulness of this technique, of course, relies on the accuracy of the initial engineering estimates.

### 4.2.1 CONTROL GROUP – COMPARISON GROUP

Regardless of the techniques used, most evaluations are based on an experimental or quasi-experimental design, in which the impact of a programme is estimated by comparing a treatment group to a comparison group**{** XE "comparison group" **}** or a control group (see also Section 5.1). The energy that the targeted population would have consumed in the absence of the programme cannot be observed directly; a proxy for this consumption must be used.

**Case Example: Energy Efficiency Check (EEC)**

Based on the new examination of the programme history we decided to perform a larger survey to ensure significant results. We chose a main sample and two comparison samples to answer our questions regarding effects related to "non-participants", self-selection, rebound, free-riders and spill-over effects, etc. In total 1,200 customers in Akershus region were interviewed.

The three sample groups were:

- Customers who participated in the EEC programme, i.e., received, completed, and returned the form;
- Customers who received the EEC, but did not complete or return the form;
- Customers who did not receive the EEC (or other EE material from the EE centre over the past year).

The survey revealed several interesting characteristics:

- About 72% of the 2,400 people that completed the EEC were men. This might indicate that men are the ones most interested in implementing EE measures in general.
- The number of people in the household does not influence the reaction to the EEC.
- The income, however, seems to make a difference in whether you use the EEC or not. People with high income are more likely to participate in the programme. It seems that the medium size households with a living area of 100-250 square metres are more likely to return the EEC.
- The heating system of the houses has little influence.
- The main reason for implementing EE measures is to save energy and reduce electricity bills. More than 40% give this answer in all sample groups. Approximately 10% want increased comfort and about 10% say that general maintenance is the main reason. Women and people in older houses are more interested in increased comfort.
- The houses in the third sample group had paradoxically implemented more EE measures that the other houses. However, the houses in this group are in general older and hence they require more maintenance and there are more young people in older houses. Both may be contributing factors to why this sample group has implemented more EE measures on the whole than those that had received the EEC.
- The reasons stated why people do not implement EE measures vary between the sample groups. In the first sample 71% said that they had already implemented the measures. In the two other samples only 50% gave the same reason. This shows again that the people who have used the EEC are already very aware of EE. Other reasons given were "no need", "new house", and "can not afford".

**Norsk Enøk og Energi, Norway**

Ideally, a control group should be used. A control group**{** XE "control group" **}** is a comparison group, which shares the important characteristics with the treatment group and either was not offered the programme or could not participate for some reason. The impact can be found by looking at the difference between the two groups. The way to achieve this is to select randomly from the same population who is to participate and who is to be part of the control group. This is, however, rarely possible in real life. Often you have to settle for a comparison group from a different population. If the comparison group is from a different geographical area, then some of the characteristics of the members will most likely be different from those of the treatment group. Another typical type of comparison is between participants and non-participants within a certain geographical area. However, this is problematic in the sense that it often is the ones most interested or with the largest potential that choose to participate.

Usually, the effect of the programme is modelled as the difference between the programme participants' consumption and that of some group that represents participant consumption in the absence of the programme, for example:

- Participant consumption before the programme may be compared to participant consumption after the programme (pre-/post-treatment design).
- Participant consumption may be compared to consumption of non-participants either after the programme or both before and after the programme (treatment/comparison group design).
- Participant consumption may be compared to consumption of a group of consumers not exposed to the programme (treatment/comparison group design).

Special issues must be addressed with each of these designs (matching of participants to non-participants, self-selection bias, accounting for programme effects on non-participants, etc.). These issues are discussed in more detail in Chapter 5.

### 4.2.2 SAMPLING TECHNIQUES

In all except the crudest impact analyses (e.g., those based exclusively on a simple engineering algorithm, its related assumptions, and no specific data from programme participants), measurements are usually taken from a *sample* of the group being analysed. Therefore, evaluations must address the reliability of using the estimate of the sample as representative of the entire population. Sampling techniques**{** XE "sampling techniques" **}** have been discussed extensively in the existing evaluation literature and generally fall into four basic categories:

- **Random samples{** XE "random samples" **}**, which select a group of consumers/installations to be analysed using a random selection process.
- **Stratified samples{** XE "stratified samples" **}**, in which the study population is broken into homogenous groups and separate samples are selected from each homogenous group.
- **Weighted samples{** XE "weighted samples" **}**, which address the possibility that some participants may have greater impacts than others.
- **Ratio estimation{** XE "ratio estimation" **}**, in which a small group is selected for detailed and often more expensive analysis (e.g., using end-use metering) and, through a linkage based on variables common to the small group and the entire population of interest, results are extrapolated to the entire population. (Sometimes, two extrapolations are done: One to a larger group analysed at a less intensive level (e.g., through survey research) and then from the larger group to the entire population.

Careful sample design is important to the evaluation of EE programmes, because often more sophisticated sample designs can allow an evaluation study to obtain more precise estimates for the same evaluation expenditure or the same level of precision at a reduced expenditure. See the references in Appendix B for more detailed discussions of sampling strategies for impact evaluations. Sampling is also discussed in many market research and non-energy programme evaluation texts.

**Case Example: Use of Electronic VSDs in Motors in the Portuguese Industry**

Electric motor driven systems account for 67% of industrial electricity consumption in Portugal. Even so, awareness of the saving potential related to installation of variable speed drives is very low. Therefore a project was developed under PEDIP II programme with the main aim to identify the energy savings potential and sensitise industrial top level decision-makers to the application of the VSD technology.

The project consisted of a pilot action, where electronic variable speed drives and soft starters were installed in various industrial plants to allow measurement of the resulting energy savings. The results were then scaled up for each industrial subsector to arrive at an estimate of the national potential for energy savings (ratio estimation). The sample of industrial enterprises selected for pilot testing of VSD technology was, however, not representative of the whole industrial sector since preference was given to the following:

- Enterprises currently employing young technicians in an energy traineeship. CCE is currently conducting an EE programme, which provides 2 months training in EE to newly educated engineers followed by a 9 months traineeship in industrial enterprises with high electricity consumption. Furthermore, the two activities are likely to strengthen one another;
- Industrial sites listed in proposals prepared by VSD technology suppliers and where the suppliers appeared willing and able to provide e.g. the data and co-operation requested by CCE;
- Sites which had the highest possible variety of equipment sizes and types within its industrial branch;
- 50% of the total equipment cost for the pilot project was financed by the government (PEDIP II Programme) and the remaining 50% by the involved industrial sites. However, the budget limit for contribution from PEDIP II was 39,904 EUR in total. Therefore a suitable mix of industries had to be construed which avoided exceeding the permitted co-financing limit of 39,904 EUR;
- The selected projects should allow testing of a great range of motive power (between 11 and 200 kW), types of equipment (drum mills, compressors, fans, etc), and types of industries (ceramics, agro-food sector, cork, textiles, and chemicals);

The distribution of pumps, fans, compressors, and other motors varies between but also within the different industrial subsectors – mainly due to differences in manufacturing processes even for similar products. Therefore, extrapolation of pilot results to a national level does not necessarily lead to trustful values.

The consequences of this was not investigated since the aim was to estimate the approximate size of the energy saving potential of VSD introduction on a national level and not to obtain exact values for each industrial subsector.

**Centro para a Conservação de Energia, Portugal**

### 4.2.3 ENGINEERING METHODS

Planners develop basic engineering algorithms to estimate programme impacts prior to programme implementation.

There are three basic types of engineering methods:

- Estimation from **simple engineering algorithms{** XE "simple engineering algorithms" **}**;

- Estimation from **enhanced engineering algorithms{** XE "enhanced engineering algorithms" **}**; and

- Estimation from **engineering simulation models**.

Even before a programme is implemented, programme planners typically develop estimates of programme impacts, using engineering algorithms. These estimates help decision-makers decide whether programmes are likely to be cost-effective and provide very rough estimates of the programmes' impacts.

For example, engineering algorithms such as the ones presented below might be the basis for programme planners' estimates of the energy (kWh) and peak load (kW) impacts for individual measures in a non-residential efficient lighting programme (seen from a customer perspective the transmission and distribution losses should be excluded):

Change in kWh = (kW of removed or replaced unit - kW of efficient unit) * Number of units * Number of hours * (100% + Transmission & distribution system loss %)

Change in peak kW = (kW of removed or replaced unit - kW of efficient unit) * Number of units * Coincidence factor * (100% + Peak transmission & distribution system loss %)

The coincidence factor is the % of units "on" during system peak period.

The number of hours may be defined **per year** or **per lifetime** of the measure.

A more sophisticated set of algorithms for the same purpose might have separate algorithms for periods thought to be significantly different from each other and might include other factors that may significantly affect the actual impact from the programme. An example of a more sophisticated set of algorithms that address "summer" programme impacts might look like the following:

Change in kWh =

(kW of removed or replaced unit - kW of efficient unit) *

Number of units *

(1 + fraction of lighting in cooled space * Additional savings for cooling reduction factor) *

Number of hours *

(1 - Free-ridership fraction + Spill-over effect) *

(1 - Rebound fraction) *

(1 + T&D system loss fraction) *

Persistence fraction.

Change in kW =

(kW of removed or replaced unit - kW of efficient unit) *

Number of units *

(1 + fraction of lighting in cooled space * Additional savings for cooling reduction factor) *

Fraction of units "on" during system peak period *

(1 - Free-ridership fraction + Spill-over effect) *

(1 - Rebound fraction) *

(1 + Peak T&D system loss fraction) *

Persistence fraction.

In developing a pre-programme estimate of programme impacts, the planners would make assumptions about the values to use for each parameter in the algorithms, based on the best available information (e.g., previous technology studies, evaluation results from other years or other utilities/countries, wild guesses, etc.).[4] If using the more sophisticated algorithms, the planners would have similar algorithms for other systematically different periods such as winter or shoulder month periods or, for load impacts, shoulder and off-peak periods of the day.

**Case Example: Energy Performance Standard in the Dutch Building Decree**

At the end of 1995, an Energy Performance Standard (EPS) was introduced in the Dutch Building Decree. The aim of this legal instrument was to reduce the energy use in new houses, but leave freedom of choice to architects, developers, and house owners regarding how they prefer to reach the required performance level.

A calculation model for determining the energy performance level of a house therefore had to be developed. A "typical" new house (i.e., average) was defined for each of the categories: Multi-family house, row house, semi-detached house, and one-family house. An "energy budget" related to the EPS level (i.e., a maximum allowed consumption) was then defined for each category. The calculation model not only included energy consumption related to space heating, cooling and ventilation, and lighting. It also included water heating to provide additional stimuli for use of residential solar hot water systems. Using the calculation model it is then the responsibility of the architect to prove that the designed house does not exceed the allowed energy budget.

At the end of 1997, a comparison of the estimated energy use (based on the EPS calculation model) and the real energy use was made. The investigation concluded that the EPS could not be used to calculate real energy consumption. It could only be used to calculate the difference between the allowed energy budget for a specific house and the estimated energy consumption prepared by the architect using the calculation model. However, in spite of the fact that the EPS model could not be utilised on an individual level, it did on a national level arrive at energy savings estimates comparable to the realised savings in 1997.

Furthermore, it was found that the impact of behaviour and the penetration of new appliances in real households had to be included in the future analyses to be able to distinguish their impact on the household energy consumption level.

**NOVEM, The Netherlands**

**Impact estimates based on planners' engineering algorithms can be improved considerably using actual programme data.**

These same algorithms could be used by evaluators to estimate the actual impacts from a programme. A range of methods could be used, from the most simple to very "enhanced" methods. Continuing our example of the non-residential lighting programme, options include the following:

- Substitute the actual number of units recorded in the programme tracking system and calculate the estimated impacts.

- Conduct telephone surveys with programme participants. Survey results could be used to:

    - Verify the number of units installed and the installed capacity (W) of both the replaced and newly installed units.

    - Obtain an estimate of all the other parameters in the algorithms (except for spill-over effects due to non-participants).

[4] *They would also make implicit assumptions about the effect of weather on certain impact parameters such as number of hours the equipment is being used.*

- Conduct on-site visits with a sample of participants large enough and selected carefully enough to permit extrapolation of results to the entire participant population or to all participants in business segments that account for the greatest percentage of expected impacts. Surveyors would conduct essentially the same survey as conducted via the telephone but, by being at the site, would also be able to more accurately estimate parameters such as number and wattage of measures installed, % of lighting in cooled space, participant spill-over, persistence (as reflected in the number of efficient units still installed) by physically observing the facility. They might also obtain a better estimate of free-ridership, either through a more rigorous free-ridership estimation technique such as conjoint analysis or because they could talk to multiple individuals that might have been involved in the decision to install the measure (e.g., facility manager, building owner, financial officer).

- Install run-time meters (e.g., light loggers) on a sample of installed efficient units, and extrapolate results to the rest of the participant population. The purpose of this approach would be to develop a more accurate estimate of hours of use and possibly, depending on the sophistication of the meter, also the actual wattage of the efficient units (instead of "installed capacity" specified by the manufacturers) and the percentage of units "on" during the system peak period. Installation of the run-time meters could occur during a site visit as described above. A second visit would be required to retrieve data from the meters, as well as the meters themselves. Of course, additional visits could also be made to collect similar data during various time periods of interest (e.g., summer, winter, shoulder months, during the business's peak production season, etc.) during which energy usage was thought to vary significantly.

- Install more sophisticated meters on a sample of installed efficient units, so that load and energy use of the installed units and of possible affected HVAC equipment can be measured, and extrapolate results to the rest of the participant population. This would permit interaction effects of lighting on heating and cooling to be more accurately estimated.

- Install run-time or more sophisticated meters, or conduct site visits prior to measure installation, so that the change in load and energy use could be estimated more accurately.

  - Both types of meters, if not installed prior to the measure installation, provide a better estimate only of post-installation load and energy usage; assumptions would still have to be made about pre-installation usage. Installing the meters prior to installation allows energy usage of both periods to be measured, so that the change in load and energy use can be measured.

  - Site visits conducted prior to measure installation might reveal that certain lights are burned out. Consequently, a comprehensive retrofit of existing lamps with high-efficiency units might yield lower savings than expected, because more lights would be operating. Pre-installation site visits might also show that hours of use prior to measure installation actually differ from that after the installation. In the absence of the pre-installation visits the evaluation team must accept the survey respondent's report of the number of non-functioning lamps and the hours of use for specific lamps. The respondent may have an interest in inflating or these numbers, or may not know at the time of a telephone survey exactly how many lamps are burned out or exactly how many hours each lamp is used.

**Engineering simulation models{** XE "engineering simulation models" **} are an engineering-based alternative to the use of simple or enhanced engineering algorithms.**

Rather than rely on the programme's engineering algorithms, the evaluation team could take an **engineering simulation** approach. Site-visit and metered data could be used to develop detailed engineering simulation models that could predict load and energy use changes resulting from measure installations. Load and energy usage of the baseline facility prototypes would be simulated, and then the simulations would be re-run using the efficient lighting measures. The programme tracking system could then be used, to extrapolate results to the rest of the participant population. Data on other impact parameters such as free-ridership, spill-over, etc., would still need to be estimated, probably using survey and verification data collected on-site.

Due to cost considerations, it is unlikely that this method would ever be seriously considered for a programme with measures as well understood and as diverse as lighting measures and for populations as diverse as the non-residential market. However, for other programmes, such as new construction or building envelope and weatherization, engineering simulation techniques provide the only real alternative to simple engineering algorithm methods.

Exhibit 4-2 sums up when it would be advantageous to use engineering methods for evaluation of DSM and EE services programmes.

*Exhibit 4-2: Rules of thumb for engineering methods.*

| When | Comment |
|---|---|
| Always (simple engineering algorithms), as a reasonableness cross-check | Simple engineering algorithms can be used, in combination with tracking system data on participants and other site-specific data (e.g., from consumer surveys) to provide a quick-and-dirty estimate of impacts. This can serve as a reasonableness cross-check on estimates produced by more sophisticated or statistical methods. |
| When both energy or load/load shape impacts must be estimated | Engineering methods tend to be the most cost-effective approach to estimating load and load shape impacts. |
| When interactions between measures or end-use equipment must be accounted for | Interactions such as impacts of improving lighting efficiency on the energy use of space conditioning equipment are difficult to account for in statistical models. Engineering models are more effective for determining the extent of these interaction effects. |
| When measures are well understood | Engineering algorithms can often be used — supplemented by tracking system and site-specific data — to estimate impacts of measures whose impacts or impact components are well understood. |
| When a wide range of heterogeneous measures are being analysed | For programmes involving multiple measures implemented by different participants (e.g., many commercial and industrial programmes), the number of participants implementing the same measure will likely be limited, making statistical methods less reliable and less precise. |

### 4.2.4 STATISTICAL METHODS

Impacts of some types of programmes can be estimated effectively using statistical analysis of energy data. Statistical approaches most often provide estimates of energy (e.g. kWh) impacts for some types of programmes as estimating load impacts with these methods requires a significant (and typically very costly) amount of participant and non-participant load data.

There are several basic statistical methods:

- Simple comparisons;
- Weather-adjusted comparisons; and

- Multivariate analyses.

The following basic evaluation issues must be addressed, regardless of the method being used:

- How did participant energy usage change after measure implementation?
- What portion of that change was due to the programme rather than other unrelated factors? Examples of unrelated factors that could be addressed include the following:
  - Changes in weather patterns from year to year;
  - Changes in disposable income, which might cause consumers to use more or less energy;
  - Changes in energy costs, which might cause consumers to use more or less energy;
  - Changes in attitudes toward energy conservation, for example, due to published reports of shortages of energy supplies or the need for environmental improvements;
  - Changes in the number or type of energy-using equipment or appliances in the home/facility;
  - Changes in the number of occupants, or the number of hours/timing**{** XE "timing" **}** of their occupancy;
  - Changes in production levels (for industrial facilities).
- What would their energy usage have been if there had been no programme?

Data on energy use is usually compiled in utility customer billing records. Energy data can also be obtained using questionnaires. This could be relevant if the programme or the evaluation is not carried out by the energy provider. Furthermore, there are programmes, which involve other energy resources than electricity and where a fuel switch is possible or desired. Also here statistical methods can be applied. For simplicity's sake the remaining part of Section 4.2 refers to **energy bill data** but the issues are also valid for other data sources.

### *SIMPLE COMPARISONS*

Simple comparisons**{** XE "simple comparisons" **}** can be made of:

- Programme participant energy usage before and after implementing programme measures. Typically, energy bills for a minimum 12-month period before measure implementation and a 12-month period after measure implementation are compared (time series analysis**{** XE "time series analysis" **}**). The difference between the two totals represents a rough approximation of the programme's gross energy impacts.
- Programme participant and comparison group energy usage after implementing programme measures. This technique uses a similar methodology as the one above, except that bills of two different groups of consumers are compared (cross sectional analysis**{** XE "cross sectional analysis" **}**), rather than bills of one group of consumers at two different times (time series analysis). This method allows the analysis to account for some exogenous (non-programme-related) causes for changes in energy use (e.g., economic conditions, weather). The method can be useful for programmes in which pre-implementation data is not

available, such as new construction programmes. However, unless the comparison group can serve as a control group for participants (shares a wide variety of characteristics with participants), there is considerable uncertainty about what participant usage would have been in the absence of the programme.

- Programme participant and comparison group energy use before and after implementing programme measures. This method combines the two above-mentioned methods.

### *WEATHER ADJUSTMENTS*

Each**{** XE "weather adjustments" **}** of the three simple comparison methods above can include an adjustment to account for the effect of weather on energy impacts. (This is of course most relevant for the time series analyses). Typically, weather data are used to provide a common basis on which to compare pre- and post-implementation billing data, by means of adjustments that account for differences in heating and/or cooling degree days**{** XE "degree days" **}**.

Weather data are also used to estimate total programme impacts over the life of the measure, through the use of adjustments to energy impacts based on how the weather associated with the billing data from the pre- and/or post-implementation periods compares to that of a typical meteorological year (i.e., compares to the most likely weather pattern to exist for the years that the measure will be producing energy impacts).

### *MULTIVARIATE METHODS*

Multivariate methods**{** XE "multivariate methods" **}** are the most complex of the statistical methods, but they also provide the greatest accuracy. There are two general types: Conditional demand models and statistically adjusted engineering models**{** XE "statistically adjusted engineering models" **}**. In both model types, regression equations are used to account for changes in energy usage due to factors unrelated to the programme (economic conditions, appliance holdings, number of occupants, production levels, etc.) and differences between the participant group and the comparison group (if a comparison group is used).

The generic form of the multivariate method is a regression analysis based on a model such as:

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + \text{etc.}$$

The coefficients $a_0$, $a_1$, $a_2$, … are determined in the regression analysis. The “y” may be the energy consumption of the individual customer. The “x”s may be background variables or characteristics (size of house, number of persons per household, average age, income, education, stock of appliances) or may describe the participation (did the customer participate in the programme or information on which measures the customer has implemented).

An important challenge in relation to using regression analysis is to select the correct model formulation. This task should not be underestimated. Often there will be an almost infinite range of possibilities. Based on a given set of data (which maybe has 10 variables) many new variables can be formed. One may for example chose to create **dummy variables{** XE "dummy variables" **}**, which each represents a certain interval of age (0-20 years, 20-30 years, 30-50 years, etc.). This can be relevant if one wishes to test whether there is a linear connection between age and energy consumption.

**Interaction variables{** XE "interaction variables" **}** may also be created by multiplication of two variables. A large number of base variables enable the creation of an even larger number of interaction variables. Interaction variables are used when the significance of $x_1$ and $x_2$ are not mutually independent, but $x_1$ maybe has a particularly large influence when $x_2$ is large. Non-linear links such as $x_1*x_1$ may also be created.

A single model with two variables

$$y = a_0 + a_1*x_1 + a_2*x_2$$

may thus turn into

$$y = a_0 + a_1*x_1 + a_2*x_2 + a_3*x_1*x_2 + a_4*x_1*x_1 + a_5*x_2*x_2 + \text{etc.}$$

A rule of thumb says that you need 10-20 times as many observations compared to the number of variables in your model. So if you have 100 observations then it will most likely not be possible to create models with more than 5 variables. The regression analysis shows, which variables ("x"s) influence significantly on the variation in "y".

Examples of regression analysis can be found in Tiedeman (1999), Torok (1999), Titus (1999) and Heinrich (1998) (fully listed in Appendix B). Togeby (2000) provides further examples on the use of regression analysis and warns against pitfalls.

Typically, two regression models are formed: One regarding the participation decision (a discrete choice model**{** XE "discrete choice model" **}**) to address the self selection issue, and one that includes a self-selection-correction variable (Inverse Mills Ratio) as part of a regression equation that disaggregates energy bills into their components, one of which is implementation of the programme measure.

The approach to deal with the issue of self-selection**{** XE "self-selection" **}** is best illustrated by an example, presented by Christie Torok et al. (1999) (the following is a quotation).

> Assume a net billing model specification that incorporates both participants and non-participants into one model. A disadvantage of this would be that the resulting sample is not randomly determined. There would be certain unobserved characteristics that influence the decision to participate. If these characteristics are not accounted for in the model, the net savings model could produce biased coefficient estimates.
>
> One solution to this problem is to include an **Inverse Mills Ratio{** XE "Inverse Mills Ratio" **}** in the model to correct for self-selection bias. This method was developed by Heckman (1976, 1979) and is used by others to address the problem of self-selection into energy programmes.
>
> It is assumed that the unobserved factors that influence participation are distributed normally. Including an Inverse Mills Ratio in the model as explanatory variable controls for the influence of the characteristics that cause participants to self-select into the retrofit programme. This corrects for the self-selection bias in the net savings regression as the unobserved factors affecting participation are now controlled for in the model. As a result, standard regression techniques should produce unbiased coefficient estimates.
>
> Goldberg and Train (1996) developed the technique of including a second Inverse Mills Ratio in the savings regression to account for the possibility that participation is correlated with the size of potential energy savings. The second Inverse Mills Ratio is interacted with a measure of energy savings, which allows the amount of net savings to vary with participation. The rationale for the second term is that those customers who have potentially large savings are more likely to participate in the programme. Consequently, the unobserved factors that are influencing participation are also affecting the amount of savings.
>
> To calculate the Inverse Mills Ratios, a probit model[5] of programme participation is estimated, and the parameters of this model are used to calculate an individual Inverse Mills Ratio for all participants and non-

[5] *A probit model is a model, which describes the probability of an occurrence. An example could be an investigation of the probability that a certain person chooses to participate in an EE Programme. If*

participants. This Inverse Mills Ratio is included in a net savings regression that combines both participants and non-participants into one model.

If the Inverse Mills Ratio controls for those unobserved factors that determine participation (i.e., the self-selection bias), and the other model assumptions are met, then the net savings model will produce unbiased estimates of net savings. The resulting statistically adjusted engineering coefficients on the energy impacts (that have been interacted with the Inverse Mills Ratios) are then used to adjust the engineering estimates of expected annual energy impacts (the original statistically adjusted engineering coefficients) for the entire participant population. This is one estimate of net ex-post energy impacts.

The latter model is a **conditional demand analysis{** XE "conditional demand analysis" **}** model. It models energy use as a function of the many end-uses of energy that exist at the consumer's home or facility, as well as any key attitudinal or other factors that may have a strong influence on energy use. Data on end-uses and other factors are obtained from surveys with those consumers whose bills will be analysed. Each period's consumption is disaggregated into its components. The difference between the pre- and post-implementation average consumption of the targeted end-use represents the programme's impact in that period (e.g. month, quarter). The differences in each period are then summed, to produce an estimate of total (e.g. annual) impacts.

The **energy usage regression equations{** XE "energy usage regression equations" **}** estimate coefficients for each end-use in the analysis, and those coefficients represent the portion of the e.g. monthly energy use that is attributable to each end-use. In a statistically adjusted engineering model, there is already a variable that includes an engineering estimate of the measure's effect on the targeted end-use. The regression estimates how that engineering estimate must be adjusted to better reflect the measure's impact on the end-use consumption. The estimated coefficient is an adjustment factor (*realisation rate*) to be applied to the engineering impact estimate.

The table below illustrates how each of the statistical methods addresses each of the key evaluation issues.

*Exhibit 4-3: How statistical methods address key evaluation issues.*

| Method | Addresses Change in Participant Energy Use | Addresses Factors Unrelated to Programme | Addresses Consumer Energy Use in Absence of Programme |
|---|---|---|---|
| Simple time-series comparison | Yes | No | No |
| Simple cross-sectional comparison | Theoretically possible sometimes | Some factors addressed | Theoretically possible sometimes |
| Simple time-series and cross-sectional comparison | Yes | Some factors addressed | Theoretically possible sometimes |
| Weather adjusted methods | - | Addresses weather effects | - |
| Conditional demand analysis | Yes | Yes, in most cases | Theoretically possible sometimes |
| Statistically adjusted | Yes | Yes, in most cases | Theoretically possible |

*background variables exist on a number of persons, who participated, and others, who chose not to, then it is possible to calibrate the probit model.*

| | | | |
|---|---|---|---|
| engineering models | | | sometimes |

As the table shows, the more advanced methods tend to better account for factors affecting energy use, which are unrelated to the programme.

All of the methods include the effects of participant spill-over and rebound, as well as first-year measure retention (as a proxy for persistence of savings), because they are based on an analysis of actual energy usage, not engineering algorithms. If participants have made additional energy-conserving improvements, unrelated to the programme (spill-over), the effects of these improvements will be reflected in their energy bills, relative to the comparison group. The same is true for additional energy use due to a perception that energy bills will be lower (rebound); this additional usage will be visible in the energy use that the statistical methods analyse.

Exhibit 4-4 shows that none of the methods is really reliable for addressing the baseline issue of free-ridership or non-participant spill-over (see Chapter 5). These analyses must be conducted separately, though free-ridership and (if non-participants are included in the analysis) non-participant spill-over data can be provided through the surveys conducted to gather the inputs for the conditional demand analysis.

*Exhibit 4-4: Rules of thumb for statistical methods.*

| When | Comment |
|---|---|
| When there is a large number of homogenous participants (e.g., broad-based residential programmes; commercial programmes, when large participant segments can be constructed) | Statistical methods are effective when a large number of observations are analysed and energy use is the only (or one of the only) factor(s) varying significantly among those observations. |
| When the audience for the evaluation demands that impact estimates be based on actual measured energy consumption | Statistical evaluations are empirical, in that the change in energy use is inferred from an examination of observed energy consumption. An estimate may also be accomplished using metered data but is much more (usually prohibitively) expensive. |
| When there are reasons to expect that consumers' behavioural responses are significant | There sometimes is a tendency for participants to operate energy-using equipment more frequently or at a higher level of intensity after participating in a programme. They may perceive that each increment of energy use costs them less, due to the energy-efficiency improvement they have made. Because statistical methods observe actual consumption, they address this phenomenon. Also, for pilot programmes (when there is a higher likelihood of non-participants representing a true control group), statistical methods can account for naturally occurring adoption of the targeted efficiency measure. Engineering methods cannot generally address either of these issues, except through assumptions. |
| When energy impacts (as opposed to load/load shape impacts) are being estimated | Statistical methods require large sample sizes, and load impact analysis using statistical methods relies on load data for a large number of consumers. Typically, this requires extensive, expensive metering, or the resulting impact estimates will be too influenced by individual (and possibly idiosyncratic) data. Such metering is typically cost prohibitive. |

# 4.3 CHOOSING BETWEEN SIMPLE AND MORE COMPLEX METHODS

A report for the U.S. National Association of Regulatory Utility Commissioners (NARUC)[6] identifies four main categories of EE measures promoted by EE programmes:

- Constant efficiency, constant load – e.g., most lighting equipment efficiency improvements (energy-efficient fluorescent lamps and ballasts, delamping) and constant load motors

- Constant efficiency, variable load – e.g., daylighting controls and energy-efficient water heaters, as well as some thermal energy storage, direct load control, and energy management systems applications; in short, measures for which the load can vary considerably, but the efficiency remains constant

- Variable efficiency, constant load – e.g., efficiency improvements to supermarket refrigerated cases, for which the temperature and humidity conditions in the store remain relatively constant but the efficiency of the refrigeration system varies considerably with outdoor temperatures and humidity (infrequent type of measure)

- Variable efficiency, variable load – e.g., heating and air conditioning measures, building envelope improvements, or variable speed motors, as well as some thermal energy storage, direct load control, and energy management systems applications for which both the efficiency and load vary with outdoor temperatures or production needs.

Such a classification of measures is helpful in that it permits statement of a basic rule of thumb for selecting impact evaluation techniques:

> *"As measures move from constant efficiency and constant load to variable efficiency and variable load, the analytic approach and data requirements become more challenging".[7]*

In other words, it is easier to predict the performance of measures or equipment when the primary components of energy use are constant. For constant efficiency/constant load measures, rather than accounting for a wide range of operating conditions, one may be able to assume just one, or a very limited number of operating conditions. Instantaneous measurements of load may suffice. For measures with variable efficiency and/or load, ongoing metering to collect data on how equipment operates at different times of day or in different seasons or in the context of different production schedules/volumes might be required.

For example, we can generally rely on the assumption that a 25 W compact fluorescent lamp that replaces a 100 W incandescent lamp will produce an energy reduction of 75 W, regardless of the time of day or time of year. We need only a simple engineering algorithm and data on the users' hours and timing of use to be able to reliably predict the energy savings[8]. In contrast,

[6] Evaluating Energy-Efficiency Programmes in a Restructured Industry Environment: A Handbook for PUC Staff *by J. Schlegel, M. Goldberg, J. Raab, R. Prahl, M. Keneipp, and D. Violette. See the appendix to this report for a summary of this document.*

[7] *Ibid.*

[8] *Average resting time would also be important in such an evaluation, because certain values of average resting time can shorten the lifetime of the CFL (about 3min.).*

effectively predicting the energy savings associated with a variable-speed motor requires a more complex analysis and additional data (e.g., load factor, and operating profile derived from end-use metering).

If the measure has been the subject of rigorous studies in the past, some information on the degree to which various components of energy use vary by user should be known. Evaluators can then make decisions about whether and which specific engineering algorithm parameters require further investigation in estimating impacts for their own programmes.

Evaluation using simple engineering algorithms, a reliable tracking system identifying programme participants, and some basic site-specific data (baseline efficiency, hours of use, etc., which often may be collected by programme implementers or via telephone interviews) may provide reasonable estimates of the programme impacts for a variety of measures. One obvious goal for sponsors of EE programmes is to find or develop a set of reliable, transferable energy savings engineering assumptions and algorithms associated with as many measures as possible that are likely to be promoted in their EE programmes.

The U.S. Environmental Protection Agency has compiled such a list as part of its Conservation Verification Protocols to be used by states wishing to claim $CO_2$ emissions reduction credits for their EE programmes as part of efforts to come into compliance with mandated emissions levels. Use of the engineering algorithms for such "stipulated measures" guarantee that a specific level of energy savings can be claimed. This savings estimate is slightly discounted to account for its imprecision and can be increased slightly if more thorough analysis is conducted by the organisation making the energy savings claim. Exhibit 4-5 presents a similar list of net-to-gross factors[9] that can be used by applicants.

Use of these assumptions and algorithms must, of course, be adjusted for any differences between the U.S. and the user country with regard to the energy use of the targeted equipment (e.g., significantly differing operating conditions or hours) and its related efficiency measure (e.g., different technical characteristics/standards for the product).

Another option for controlling evaluation costs is to conduct joint evaluation research, several sponsors funding a rigorous study. In this way the costs of investigating less well-understood, but important, EE measures, or specific impact parameters for such measures, can be shared. Future evaluations by the individual sponsors can then rely on estimated parameters from the joint study with confidence, without having to conduct costly evaluations on their own.

[9] *See the following section for definition. Please note, that the net-to-gross factor refers to the relation (net/gross) and not the sequence in reaching the figures (first calculating the gross impact, then applying the factors to arrive at the net impact).*

*Exhibit 4-5: Default net-to-gross factors.*

| Conservation Measure | Net-to-Gross Factor |
|---|---|
| Refrigerators | |
| - Pick-up | 0.70 |
| - High-efficiency replacement | 0.90 |
| Residential Water Heating Measures | |
| - Insulation blankets | 0.60 |
| - Anti-convection valves | 0.90 |
| - Pipe Insulation | 0.60 |
| - Low-flow showerheads and faucet aerators (utility-installed) | 0.70 |
| - Low-flow showerheads and faucet aerators (customer installed) | 0.50 |
| - Heat pump water heaters | 0.95 |
| Ground Source Heat Pumps for Homes | 0.95 |
| Higher Efficiency Lighting in Office Buildings | 0.60 |
| De-lamping in Commercial Buildings | 0.80 |
| Exit Sign Light Replacements | 0.60 |
| Higher Efficiency Street Lights | 0.90 |
| Higher Efficiency Motors for Constant Load Applications | 0.60 |

*Source: The User's guide to the Conservation Verification Protocols, Version 2.0, United States Environmental Protection Agency, Air and Radiation (6204-J), EPA 430-B-96-002, April 1996.*

## 4.4 ECONOMIC EVALUATION

One{ XE "economic evaluation" } of the central reasons for performing evaluation research may be to determine the cost-effectiveness of the programme. Do the programme benefits outweigh costs?

### 4.4.1 GENERAL RECOMMENDATIONS

The economics surrounding the evaluation must be considered. Some highlights are according to the Swedish Evaluation Guidebook, NUTEK et. al., 1993:

- The approach used to carry out the ex-post evaluation must be similar to that applied before initiation of the EE programme (i.e., that used for ex-ante evaluation) to allow easy comparison.

- Programme cost accounting should include all programme-related costs at market value and should be used for evaluation of most energy services (i.e., internal company transactions should booked at market value). Care should be made to include all "sunk" programme costs.

- Only variable costs should be included, i.e., costs incurred as a result of the programme. Fixed costs such as non-programme related administration costs should be omitted.

- It is recommended to use net-present-value calculations to allow taking the time factor into account (occurrence in time of payments and disbursements) especially for programmes intended at achieving long-term impacts. The important point is to take into account the long-term consequences of the programme and the cost of capital (exclusive of tax). Note that when considering tax impacts, the programme should not be considered independently but as a part of a larger enterprise. Furthermore, it is recommended that all financing costs (such as tax-deductible interest on capital debt) be taken into consideration when estimating the cost of capital.

- No generally accepted norm exists for adjusting for inflation, which is why most companies refrain from doing so. In this situation the costs are underestimated; inflation, in most cases, affects revenues more than costs because they occur later in the programme life.

- And finally, when evaluating the programme, evaluation of the consequences of alternatives can also be relevant.

### 4.4.2 BENEFIT/COST ANALYSIS

A publication prepared for the European Commission addressed benefit/cost analysis of EE programmes in detail[10].

Exhibit 4-6 shows the basic framework for B/C analysis{ XE "B/C analysis" }, including costs, benefits, benefit/cost ratio and other impacts from each of six perspectives.

Both costs and benefits can be indicated in present values{ XE "present values" } or by the balance (benefits minus costs, i.e., the net present value (NPV)). Calculating the B/C ratio{ XE "B/C ratio" } (the benefits divided by the costs) is not a necessary step, but enables a quick glance comparison of a variety of possible programmes.

*Exhibit 4-6: Framework for B/C analysis.*

| Perspective | Costs | Benefits | B/C Ratio | Other Impacts |
|---|---|---|---|---|
| Customer | | | | |
| Distribution | | | | |
| Wholesale Utility | | | | |
| Government | | | | |
| Society | | | | |

The following Exhibit 4-7 shows the factors included in the “primary equation” (under “costs” and “benefits” in the table above) as well as factors that are “otherwise accounted” (under “other impacts” above).

[10] *The following consists of partial quotes from the report* European B/C Analysis Methodology: A guidebook for B/C Evaluation of DSM and Energy Efficiency Services Programmes, *prepared for the European Commission (DG 17), by SRC International ApS (Denmark) and a project advisory committee with representatives from numerous EU countries, February 1996. This document is summarised in the appendix.*

Monetized costs and benefits are typically given the most weight in the benefit/cost analysis. Other impacts, however, are sometimes critical to decision-making, and they are included formally in the matrix so that they can be included if desired. The following description, from the guidebook, presents a rationale for the consideration of quantitative and qualitative analyses.

*Exhibit 4-7: Overview of relevant benefits and costs by perspective.*

| Perspective | Included In Primary Equation | Otherwise Accounted |
|---|---|---|
| **Participating Customer** | Consumption of Other Fuels<br>Change in Energy Bill<br>Industrial Productivity<br>Customer Capital Investment<br>Customer O&M<br>Utility Incentives<br>Third Party Incentives<br>Tax Credits<br>Taxes<br>Other Customer Transaction Costs (*)<br>Customer Value (*)<br>Tariff Changes (*) | Proven Performance<br>Ease of Implementation<br>Availability of Capital<br>(Other Customer Transaction Cost (*))<br>(Customer Value (*))<br>(Tariff Changes (*)) |
| **Non-participating Customer** | Tariff Changes (*) | (Tariff Changes (*)) |
| **Generation and Transmission Utility** | Energy Generation Costs<br>Generation Capacity Cost<br>Transmission Capacity Cost<br>Power Purchase Revenue<br>Wholesale Utility Programme Costs<br>Wholesale Utility Incentive Payments<br>Risk and Reliability (*) | Public Image<br>(Risk and Reliability (*)) |
| **Distribution and Supply Utility** | Power Purchase Cost<br>Utility Revenue Change<br>Distribution Capacity Cost<br>Distribution Utility Programme Costs<br>Distribution Utility Incentive Payments<br>Tariff Changes (*) | Market Share<br>Public Image<br>Proven Performance<br>Ease of Implementation<br>Ease of Evaluation<br>Availability of Capital<br>Cash Flow<br>(Tariff Changes(*)) |
| **Government** | Tax Revenues<br>Government Programme Costs<br>Tax Credits<br>Environmental Effects of Supply (*)<br>Environmental Effects of Consump. (*) | Industrial Productivity<br>Regional Employment<br>Public Image<br>Diminishment of Natural Resources<br>Anti-Competitiveness<br>(Environmental Effects of Supply (*))<br>(Environmental Effects of Consump. (*)) |
| **Society** | Energy Generation Costs<br>Generation Capacity Cost<br>Transmission Capacity Cost<br>Distribution Capacity Cost<br>Utility Programme Costs<br>Government Programme Costs<br>Third Party Programme Cost<br>Customer Capital Investment<br>Customer O&M<br>Environmental Effects of Supply (*)<br>Environmental Effects of Consump. (*)<br>Tariff Changes (*)<br>Other Customer Transaction Costs (*)<br>Customer Value (*) | Industrial Productivity<br>Regional Employment<br>Diminishment of Natural Resources<br>Anti-Competitiveness<br>(Environmental Effects of Supply (*))<br>(Environmental Effects of Consump. (*))<br>(Tariff Changes (*))<br>(Other Customer Transaction Costs (*))<br>(Customer Value (*)) |

*(*) - Include if translated into monetary equivalents.*

### 4.4.3 QUANTITATIVE METHODS

Quantitative methods{ XE "quantitative methods" } seek to measure all benefits and costs into a monetary unit. Recognising that numerous costs and benefits naturally are measured in different units, a dominant task in the quantitative task is to convert all relevant benefits and costs into a monetary value. If this task can be satisfactorily solved, the remaining task is methodological, e.g., calculating a benefit/cost ratio for measuring cost-effectiveness of DSM and energy service programmes.

Cost-effective outcomes may be expressed either as having a positive net present value (NPV) or having a benefit/cost ratio (BCR) in excess of one. In both cases the basic idea is simple; a programme is cost-effective if and only if benefits outweigh costs. Formally, the equivalence between NPV and BCR can be expressed as:

$$NPV = B - C \quad \text{or} \quad BCR = B/C$$

where

| | | |
|---|---|---|
| B | = | Present value of total benefits; |
| C | = | Present value of total costs; |
| NPV | = | Net present value; |
| BCR | = | Benefit/cost ratio. |

There are a number of factors in favour of using a quantitative approach. These include:

- It is consistent – A monetary value of relevant benefit and costs provides consistent ranking of DSM and energy service programmes.
- It is transparent – It provides political and public insight into preferences and value trade-offs.
- It is flexible – Monetized benefits and costs are flexible, allowing for implementation into other planning contexts, and for application to other sectors of the economy than the energy sector.

The major difficulty in using a quantitative approach is establishing monetary values for the benefits and costs, some of which may appear as unquantifiable. Some benefits and costs are difficult to measure in a physical unit, and even more difficult to monetize. For example, reliability of an electricity supply system can be represented by the cost of unserved supply, or environmental impacts may be represented with the costs of environmental damage. The general pitfalls when using quantitative methods can be summarised as follows:

- Boundaries – There may be inconsistent boundaries between what should be included as part of the cost or benefit and what should not.
- Data – Sufficient reliable data may be difficult to establish.
- Illusory precision – A monetary value may imply more confidence in the accuracy of its value than warranted, considering the great uncertainty involved.
- What counts, versus what is countable – It is often easy to mistake what is countable for what really counts, thereby omitting important factors because they appear as non-quantifiable, even though they may be of crucial importance.

Another consideration with the measurement of quantitative impacts is whether full benefits and full costs should be used as opposed to *incremental* benefits or costs. There is no single rule that can be used to determine which is most appropriate. Generally, incremental impacts are used, but there are many cases in which totals are used. In these cases it is important to explain that only incremental impacts are affecting the BCR (but not the NPV).

### 4.4.4 QUALITATIVE METHODS

The**{** XE "qualitative methods" **}** principal motivation for using a qualitative approach relates to the above stated shortcomings of a quantitative approach. Basically, these problems relate to **the problem of monetizing relevant attributes**. How is it possible to establish representative monetary values that are acceptable to all concerned groups and individuals?

The qualitative approach recognises that there are relevant benefits and costs from DSM and energy service programmes that (1) cannot easily be measured in a monetary value, and (2) are non-commensurable or non-comparable. The perception of benefits and costs will vary among decision-makers, and the trade-off between benefits and costs will vary. In a qualitative approach, all relevant benefits and costs keep their original units, and the evaluation becomes the qualitative task of trading off benefits and costs to find the best solution, or rather the group of solutions that contain the best choice. The selection process is performed in the presence of irreducible uncertainties.

Because the numerous costs and benefits may be measured in different units, the decision-maker must be able to make trade-offs among the values. This can be difficult and often depends on qualitative judgements that may differ among decision-makers. The desire to make these trade-offs more explicit, is the reason that quantitative methods are often used that require monetization of all impacts.[11]

For a more detailed discussion of the EU benefit/cost analysis method, see the summary of this document in Appendix B.

[11] *Ibid.*

# 5 KEY IMPACT EVALUATION CONCEPTS

Before entering a discussion on choosing evaluation approaches and procedures, some key concepts must be defined and explained to give an impression of the associated effort and cost requirements.

## 5.1 GROSS & NET PROGRAMME IMPACT ESTIMATION

All programme impact evaluation involves comparing what happened in the context of the programme to what would have happened in the absence of the programme. In ideal circumstances, this involves comparing a **treatment group{** XE "treatment group" **}** (those exposed to the programme) to a control group (those not exposed to the programme) with regard to the characteristic of interest (improvement in health, educational performance, change in energy use, etc.). Ideally, the **control group** is identical to the treatment group in every way other than having received the treatment. The performance of the control group represents the **baseline** against which the performance of the treatment group is compared. The baseline thus represents the performance of the treatment group in the absence of the programme.

### 5.1.1 GROSS IMPACT ESTIMATES

EE programme planners typically develop an estimate of the impacts that will result from a programme *prior to the programme being implemented*. In this way, programme costs can be compared to programme benefits, to ensure that (1) it is more beneficial (cost-effective) to implement this programme rather than another and (2) the programme benefits will outweigh programme costs.

To accomplish this task, planners must make assumptions about **programme baselines{** XE "baseline" **}** i.e., about energy use (and for some programmes initial levels for key market indicators) in the absence of the programme. These assumptions in turn are based on assumptions about the components of energy use, such as equipment/facility efficiency levels, hours of use of the targeted technology, technological development, market saturation, etc. The "assumed baseline" functions as the planners' estimate of the performance of an ideal control group for the programme.

**Case Example: Campaign for Lower Clothes Washing Temperatures**

The surveys showed a remarkable decrease in the frequency of washes at 90°C or more and that the general shift was from 90°C towards 40°C washes. It is difficult to conclude what part of the change is caused by the campaign. A background trend towards reducing the frequency of 90°C washes does exist. When a trend already exist it is imprecise to use the start year as the baseline.

German data indicate a trend towards reducing the number of 90°C washes by 1% per year. During the campaign the reduction was 3% per year in Denmark. However, it is close to impossible to determine which cultural differences exist between Denmark and Germany concerning washing habits.

**Elkraft System, Denmark**

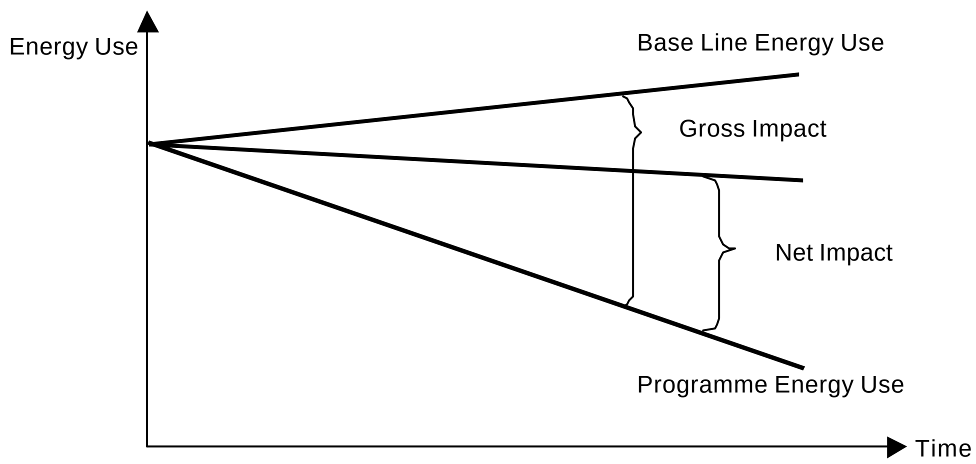
The difference between the assumed baseline energy use and the programme energy use is the "**gross impact{** XE "gross impact" **}**" of the programme.

Historically, planners and evaluators have had a great deal of difficulty estimating the programme baseline. In most EE programmes, everyone within the jurisdiction of the programme sponsor is typically exposed to the programme and so a real control group does not exist. Everyone has been exposed to the programme, and some portion of those who have been exposed become programme participants. Those who do not become participants may be categorically different from participants and so cannot serve as a control group.

This non-participant population may include many who have actively rejected the measure or who do not need what the programme has to offer. There is also self-selection bias (see also Section 4.2). Many consumers who would make the targeted improvements even in the absence of the programme (free-riders) are likely to become programme participants if a programme exists, leaving fewer of this type of consumer in the non-participant population. As a result, the average consumption of the non-participant group is likely to be higher than it would have been if these free-riders were still in the group. The non-participants' energy use is certainly not a good proxy for participant energy use in the absence of the programme. Therefore, an estimate of the programme baseline must be developed in some other manner.

This is one of the major challenges in evaluating EE programmes – defining the baseline

*Exhibit 5-1: Graphic representation of the relation between baseline, gross impact, and net impact.*




## 5.1.2 NET IMPACT ESTIMATES

While planners (and evaluators) can certainly do much to maximise the accuracy of their impact estimates, much is out of their control. Causes for inaccuracy in impact estimates are numerous and might include factors such as the following:

- Equipment nameplate efficiency[12] is different from its real-world efficiency.

*[12] The efficiency or typical energy use of a product, as reported by the manufacturer.*

- Planners and evaluators were not aware of existing plans of customers to install efficient equipment without an EE programme.

- Consumers/businesses studied prior to the programme are different from those who actually participate in the programme – the households may be larger or smaller; the equipment installed may have a larger or smaller energy capacity than expected; etc.

- Consumer/business or trade ally attitudes and practices may change during the programme for reasons having nothing to do with the programme (e.g., changes in economic conditions or employment, or a major environmental incident may occur that suddenly creates a strong interest in energy efficiency).

- Consumers/businesses use the targeted energy-using equipment differently than expected (e.g., they may use it more hours per day or less hours per day; they may use it only at partial capacity).

In the context of EE programmes, **net-to-gross{** XE "net-to-gross" **} estimation** differentiates between total changes in energy use (gross impacts), on the one hand, and only those changes that were specifically caused by the programme (**net impacts{** XE "net impact" **}**), on the other. The net impact is estimated by applying **adjustment factors{** XE "adjustment factors" **}** to the estimated gross impact. Such adjustment factors include free-ridership, spill-over, persistence, and rebound effects.

An example of an energy savings engineering algorithm which acknowledges net-to-gross factors**{** XE "net-to-gross factors" **}** (in italics) follows:

Change in kWh = (Replaced load - Programme load) *
Number of hours of use/year *
Number of measures *
Lifetime of measure *
*(1 - Free-Ridership Fraction + Spill-over Fraction) **
*(1 - Rebound Fraction) **
*(Persistence Fraction)*

An important element of impact evaluation is to provide feedback to programme planners on the size of these adjustment factors and their significance in relation to the specific programme. These factors are also referred to as “net-to-gross adjustment factors” since they describe the size of the net impact relative to the gross impact (net/gross).

If decision-makers could count on planning estimates and evaluation estimates being the same, they would not need impact evaluation activities, to know how the programme affected energy use (though evaluation could contribute to an understanding of customer satisfaction, how to improve programme cost-effectiveness, and how the programme is affecting the market). Of course, this is rarely the case, and sometimes the difference between the two impact estimates is quite large.

Furthermore, if decision-makers cannot obtain a reasonable estimate of what would have occurred in the absence of the programme (i.e., a credible baseline), they can never be sure of whether they are wasting money on trying to persuade consumers and businesses to take actions that they already plan to take or that they would take even without any persuasion. This not only affects whether a programme is cost-effective enough to continue; it may also affect decisions about implementing one programme rather than another. One programme may provide a much greater gross impact than another, but that other programme may actually yield a greater net impact.

Each of the above-mentioned net-to-gross adjustment factors is described briefly below.

# 5.2 NET-TO-GROSS ADJUSTMENT FACTORS

The impact of programmes may be adjusted analysing four different factors: Free-ridership, spill-over, rebound, and persistence of savings.

## 5.2.1 FREE-RIDERSHIP

Free-ridership**{** XE "Free-ridership" **}** is that portion of gross programme impacts that would have occurred even if there had been no programme. A free-rider is a customer who would have adopted the actions recommended by the programme even without the programme and who participates directly in the programme. Because some free-riders may have larger end-use equipment or may use it more than others, they may have different levels of free-ridership impacts that must be deducted from the gross impact estimate to obtain the net programme impact estimate.

For example, assume a programme had two participants yielding a gross impact/saving of 300 kWh/year. Also assume that Participant A installed equipment that was twice as energy-intensive as that of Participant B, but both pieces of equipment had the same EE rating. If Participant A is a free-rider, one cannot simply say that free-ridership is 50% and arrive at a net impact estimate of 150 kWh/year. One must deduct from the gross impacts that portion of the impacts contributed by the free-rider – in this case, two thirds of the impacts – and reduce the net impact estimate to 100 kWh/year.

Free-ridership is of three types:

- Pure or full free-ridership,
- Partial free-ridership,
- Deferred free-ridership.

**Pure or full free-ridership** exists when all of the gross impact related to an installation or some other unit of programme implementation would have occurred exactly as it did in the programme, even if the programme had not existed.

For example, assume that the planning estimate for the programme's baseline is that firms installing new boilers will install boilers having 84% efficiency, and that their existing boilers are typically 77% efficient. An industrial firm may have planned to install a 90% efficient boiler prior to hearing about a programme promoting the installation of energy efficient boilers. The firm participates in the programme – in order to receive a programme incentive that might be offered, or publicity, or special financing – and installs a 90% efficient boiler. All of the energy savings associated with the change between a 90% efficient boiler and an 84% efficient one would have occurred if there had been no programme. Therefore, all of these savings should be discounted from the estimate of the programme's gross energy savings, to accurately portray the true level of (net) energy impact the programme had. This programme had no energy impact on this participant.

**Partial free-ridership** exists when only some portion of the gross impact**{** XE "gross impact" **}** would have occurred in the absence of the programme.

For example, the same firm may have planned to install the 90% efficient boiler but, due to the programme incentives or information, installs a 92% efficient boiler. The gross savings represented by the difference between 92% and 84% efficiency must be partially discounted, to account for what would have happened in the absence of the programme. Rather than an 8% difference in efficiency (92%-84%), the programme is responsible only for a 2% difference (92%-90%) and the energy savings associated with that difference.

**Deferred free-ridership** is more complex. It exists when some portion of the gross impact would have occurred in the absence of the programme, but would have occurred at a later date.

An industrial firm installs a 90% efficient boiler but reports that it had planned to install a more efficient boiler, anyway, in two more years. The evaluation must determine (1) how likely this later installation would have been (in light of the many factors that might influence the firm's purchase decisions in the next two years), and (2) what the efficiency of that more efficient boiler would have been. If the firm's stated intention is taken at face value, the gross savings estimate must be changed to reflect the fact that (1) the baseline is not 84% (the efficiency of new boilers firms would install in the absence of the programme) but rather 77% (the efficiency of the existing boiler), and (2) the time period over which the difference between the new baseline energy use and the programme energy use exists is now only two years rather than the life of the boiler.

Many evaluators address this problem by applying a **greater discount factor** to the free-ridership estimate the further into the future the planned installation is. For example, if the firm reports that it planned to purchase a 90% efficient boiler in two years time, the evaluator could:

- Calculate the savings associated with switching from a 77% efficient boiler to a 90% efficient one, for two years.
- Calculate the savings associated with switching from an 84% efficient unit (if the baseline is assumed to be constant for the next two years) to a 90% unit, for the life of the measure minus two years.
- Develop a discounting factor to reflect the probability that the firm's estimate two years into the future is incorrect; rather than discounting all the savings associated with the 84% to 90% efficiency change-out, only 75% or 80% of those savings might be discounted.
- Or, instead, the evaluation could take a simpler approach, either assuming that this firm is contributing no net savings (ignoring the two-year delay and labelling the savings as pure free-ridership) or – applying a discount to the pure free-ridership to account for the probability that the firm's estimate two years into the future is incorrect – count as net savings 20-25% of the difference in savings associated with the switch from a 90% efficient unit and an 84% efficient unit.

Several **methods** are at evaluators' disposal in estimating free-ridership.

One is to conduct surveys/interviews with participants and ask them what they would have done in the absence of the programme. This method has been used frequently in the past and is often used along with more sophisticated methods, as a cross-check. The problems with this method include the following:

- First and foremost, many participants simply do not know what they would have done in the absence of the programme. They may be able to guess but they really may not know.

- Those responding to survey or interview questions may have a tendency to say what they think the interviewer wants to hear (halo effect), or what they think will make them appear more knowledgeable, more environmentally conscious, etc. (self-agrandissement).

- Those responding may not remember their state of mind and intentions at the time of their purchase decision.

- Survey respondents may also fear that certain responses will jeopardise their eligibility to participate in the programme, participate in future programmes, receive any programme incentives that may be offered, etc.

To **maximise the validity** of the survey results evaluators may take some or all of the following steps:

- Conduct the survey as soon as possible after the participation decision.

- Assure the participant that no programme benefits will be jeopardised by how the respondent answers the survey questions. Inform the participant that the responses will be kept strictly confidential and reported only in aggregate.

- Rather than asking one question about intentions, ask a series of questions that require the participant to demonstrate in different ways that he/she would have taken the same action in the absence of the programme. For example:

    - Ask whether the participant has taken other efficient actions in the past. Free-ridership would be more likely if the participant has taken such actions.

    - Ask why the participant participated in the programme. Free-ridership would be implied if participants participated to receive a programme incentive or recognition, but not if they participated to save energy or reduce their energy bills; they could have done that even without the programme.

    - Ask why the participant installed the targeted equipment. Free-ridership would be implied if they installed the equipment strictly to reduce their energy bills, but not if they installed it in order to obtain a programme incentive or recognition through the programme.

    - Ask whether the participant had prior plans to install this type of equipment, whether these plans were for a specific efficiency level, what that efficiency level was, and how much more the participant expected to pay for the more efficient unit than the standard replacement unit. Free-ridership would be implied if the participant's plans were specific and the participant was fully cognisant of the cost premium of the more efficient equipment.

    - Provide typical costs for more efficient units (the efficiency level purchased, as well as units at higher efficiency levels) and for a standard unit, and ask participants how likely they would have been to pay the incremental cost for each if there had been no programme. Free-ridership would be implied if the they would have purchased units of the same or greater efficiency they purchased under the programme.

Another possibility is to conduct research with a **quasi-control population**, such as one in a different region or country. The difficulty here, of course, is that of making a convincing case that this "control" population is similar enough to the population that was exposed to the programme to really represent what the participants would have done in the absence of the programme. Unfortunately, populations in different geographic or political regions can exhibit a wide variety of differences that may affect their behaviour regarding the use of energy and purchase or installation of energy-using equipment, such as:

- The types of products offered by retailers.
- Energy prices.
- General economic conditions and employment.
- Energy-related infrastructure.
- Product distribution channels.
- Common trade ally practices.
- Awareness/understanding of energy efficiency.
- Awareness/understanding of the product or practice being promoted by the programme.
- Other cultural differences.

Thirdly, the evaluator may use **conjoint analysis{** XE "conjoint analysis" **} or discrete choice modelling** to identify how participants would have behaved in the absence of the programme. Both methods have the advantage of obtaining information on participant behaviour indirectly, so that the participant cannot try to portray themselves in a certain light (i.e., the methods address halo effect and self-agrandissement). Both methods have been used extensively in market research, mostly to gauge the future actions of consumers rather than to model their past actions.

### 5.2.2 SPILL-OVER

Spill-over**{** XE "spill-over" **}** can be defined as energy impacts caused by the programme other than those resulting from participants making the specific improvements targeted by the programme.

The most frequently cited examples of spill-over in EE programmes include the following:

- Participants are sometimes influenced by the programme to make EE improvements not directly targeted by the programme, perhaps due to an increased awareness of the benefits of energy efficiency in general.
- Consumers make the efficiency improvements promoted by the programme because of the programme, but do not bother to officially participate or let the programme sponsor know they are making these improvements.
- Trade allies are influenced by the programme to change what they recommend to their customers or change the types of equipment they stock because of the programme. Some

portion of the consumers with which they interact are affected by these actions to improve their energy efficiency but may never even know of the programme.

In each of these cases the programme is responsible for impacts outside of the formal programme participation process. **Methods for estimating spill-over** generally include the following:

- Surveys of non-participating consumers – These are subject to the same type of problems as those affecting free-ridership surveys and are generally addressed in similar manner. In fact, evaluators must be careful to differentiate spill-over effects from instances in which programme non-participants would have taken the action even in the absence of the programme.

- Surveys with trade allies and analyses of sales data and stocking practices – Depending upon the trade ally group being examined, such data may be very difficult to collect: It may be considered proprietary information, or it may not be readily available from the trade allies in a form which can be used for the evaluation. If spill-over effects are thought likely in a programme, it is best to try to plan trade ally data collection activities into the basic design of the programme (e.g., payment of trade allies to collect data, or providing trade allies with forms ahead of time on which to record the data).

### 5.2.3 REBOUND

Rebound**{** XE "rebound" **}** is increased energy use caused by participants **trading** some portion of their programme induced energy "savings" for other benefits.

Rebound can manifest itself in several ways, depending on the market sector targeted by the programme. In residential programmes, some participants may feel that since they are saving money on their energy bills as a result of making the EE improvement promoted by the programme, they can afford to use more energy. They may end up having higher energy bills, unchanged energy bills or even lower energy bills, but they have traded their energy savings for some other benefit, typically convenience or comfort.

In commercial/industrial programmes, the programme may make it possible for some participants to produce or sell additional products at no additional energy cost. For example, they may be able to afford to keep a retail store open longer, or add or extend a work shift for a production process, because per-hour or per-unit-of-production energy costs have decreased.

Survey research is typically used to identify such changes in participant energy use. For some programmes, pre- and post-installation metering of the affected equipment may be worthwhile, especially if estimated savings from a single installation is relatively high.

### 5.2.4 PERSISTENCE OF SAVINGS

Persistence of savings**{** XE "persistence of savings" **}** is the ratio between the energy use associated with programme participation and the energy use baseline which continues throughout the life of the EE measure, measured in percent.

Some programme participants remove or never install the more efficient equipment promoted by the programme, for a wide variety of reasons (lack of time, complexity of installation, aesthetics, unsatisfactory performance, impracticality), or they discontinue the energy efficient behaviour because it becomes impractical, makes them too conspicuous, etc. The persistence issue is an important one for evaluators because lack of persistence can have very significant effects on overall net programme savings estimates. For example, if an EE measure with a 15-year lifetime is removed after only two years, most of the savings thought to result from that installation will not materialise.

Programmes with significantly low persistence rates may be redesigned, to minimise this problem, through better programme targeting, more comprehensive programme promotional materials, or through obtaining commitments from participants about persisting with the programme measure or activity.

Persistence of energy savings has three components:

- Measure retention – Is the EE measure still installed?
- Effective measure life – How long does the measure continue to function at its rated efficiency?
- Rate of technical degradation of performance – At what rate does the technical performance of the measure degrade?

Both the effective measure life and the rate of technical degradation of performance only matter to the extent that they differ from the "replaced load", i.e., the equipment or behaviour the measure is replacing. If both the baseline/replaced load and the programme load have the same measure life and the same rate of technical degradation, these factors can be ignored. "Measure retention" is usually more of an issue. EE measures may be removed for many reasons, as noted above.

Some persistence research may be based on existing data from manufacturers and other sources. Literature from product testing laboratories and/or manufacturers can be examined, to determine whether the effective lifetime and the rate of **technical degradation** of the programme measure differs from that of the technology or activity it is replacing. This provides an estimate of the energy savings that are technically possible.

**Measure retention research** typically occurs both at the same time as the rest of the evaluation and also at later dates. For some EE measures, telephone surveys can be conducted with participants, asking them whether measures are installed and whether they are still functioning. More rigorous approaches use on-site surveys, so that the researchers can actually observe that the measure is still in place and functioning properly. On-site surveys are usually recommended, especially for measures involving multiple installations at one site. Surveys are typically conducted with a carefully constructed sample of participants, and the results are generalised to the entire population, accounting for any segment differences.[13]

A more rigorous method for estimating measure retention is survival analysis. Survival analysis is a technique used in bio-statistics, typically to estimate human life expectancies. In the EE programme application of this technique, measure failure/removal is tracked for a specific programme-year population, and a survival function is estimated to predict the distribution of measure failures/removals over the life of the measure. Of course, the more years of failure/removal rate data one has, the better the function can be estimated. This technique requires that numerous measure retention surveys be conducted, to provide data for the model being developed. It can therefore be used in combination with the measure retention technique.

[13] *Measure retention rates can vary by business type, housing type, socio-economic status, etc.*

However, decision-makers may not be willing to conduct numerous retention studies unless there is strong reason to believe that measures will fail or be removed at a significant rate, in which case they may reconsider programme implementation entirely or at least a significant programme redesign.

However, measure retention studies only result in an indirect estimate of the persistence of the energy savings resulting from a programme. We are interested not in whether the measures are still installed, but in whether they are still yielding energy savings.

**Billing analysis** may be used to estimate persistence of savings for certain types of EE measures, and with certain caveats. The most likely chance of success with persistence billing analysis study is when it is used to estimate the persistence for certain residential EE measures.

Billing analysis reveals energy savings if they represent a significant percentage of energy use (e.g., ≥10%). Applying this method to measures yielding smaller savings percentages is generally not feasible because the small savings tend to be masked by the imprecision of the impact estimate.

A large participant population is also necessary. The **size of a participant population** tends to diminish over time. Businesses close, move, or expand production. Households move, change their appliance mix, change their demographics (higher or lower income, greater or fewer occupants, changes in the age of occupants and therefore in their energy use). These factors result is a significant whittling away of the available sample frame for later persistence studies.

In addition, a multitude of factors unrelated to the programme can affect participant energy usage in the years between original participation and later, follow-up persistence studies. These factors confound efforts by evaluators to attribute differences in energy use solely to the original programme. Measuring persistence using billing analyses is both risky and expensive.

# 6 SELECTING IMPACT EVALUATION STRATEGIES

The appropriateness of a given evaluation strategy will depend on the type of programme or project subject to evaluation. In this chapter guidelines for selecting an impact evaluation strategy are given for six different types of programmes/projects, namely

- Targeted information programmes,
- Market transformation programmes,
- Transmission and distribution programmes,
- Load management programmes,
- Customer retention programmes, and
- Energy service company projects.

## 6.1 TARGETED INFORMATION PROGRAMMES

Most evaluation in the EE field has been conducted for incentive programmes, which count as participants only those consumers who implement the targeted programme measures. Information programmes, in contrast, attempt to influence measure implementation decisions by educating the consumer or business so that they will then implement measures at a later date. The implementation effect is thus one step removed from the programme.

The main problem of evaluating information programmes is that it can be very difficult to assess which piece of information made the target group react. A further complication is that the full programme impact may not be immediately visible – there may be a time delay in the impact (e.g. the full effect of educating school children in energy efficient behaviour may not be seen until years later). How can programme expenses be justified when impact cannot be directly measured or proved? Can a greater uncertainty in proof of impact be accepted? How is such a programme optimised? This “lack of proof” is why the majority of information programmes are carried out by public authorities or on behalf of these as public service obligations.

### 6.1.1 TYPES OF INFORMATION PROGRAMMES

From an evaluation point of view, there are two basic types of information programmes**{** XE "information programme" **}**: Those for which the participant is **known** and those for which the participant is **not known**. This section deals with targeted information programmes for which the participant is known, such as energy audits, or training seminars and workshops. Programmes with unknown participants may be treated similarly to market transformation programmes, with regard to evaluation. Evaluation of these programmes is addressed in Section 6.2.

For targeted information programmes for which the participant is known, such as energy audits, training seminars, or workshops, several characteristics facilitate the evaluation effort. This is very clear in energy audit programmes:

- Contact information for each participant is recorded in the programme tracking system.

- Engineering algorithms are used to generate impact estimates for each recommended measure for each participant, as part of the tracking system data.

- Well-designed programmes include a statement by each participant, recorded in the tracking system, regarding whether they plan or are likely to make specific EE improvements.

**Case Example: Campaign for Lower Clothes Washing Temperatures**

At the campaign's outset in 1995, electricity consumption for washing and drying clothes accounted for approximately 18% of the Danish households' total electricity consumption. Washing alone accounted for 4.5%. Washing at 90°C or more accounted for 15% of all washing in 1997 - a high percentage in comparison with other European countries. This combined with the fact that washing at 90°C uses approximately twice as much electricity as washing at 60°C, and modern detergents make washing at temperatures above 60°C superfluous, motivated the campaign.

It was also estimated that the biggest obstacle towards changing the washing habits in the target-group was objections that the clothes would not be completely clean, odours would not be removed or washing at lower temperatures is unhygienic. As background and foundation for the campaign the National Consumer Agency therefore made a study together with the Danish Technological Institute, which showed that there were no health or hygienic problems connected with washing household clothes at only 60°C.

The aim and message of the campaign was that one could lower the washing temperature, and thereby improve the environment and save electricity, without lowering the cleanliness of the clothes or comfort of the consumers.

**Elkraft System, Denmark**

### 6.1.2 DETERMINATION OF IMPLEMENTED MEASURES

The major task for the evaluation team is to determine, which measures were implemented by the participant. This can be done as follows:

- Select a sample for verification surveys, which best represents the participation population, in terms of total programme savings. (Total programme savings here means total savings associated with measures participants reported they planned to implement or were likely to implement.) Especially for non-residential audit programmes, a sample stratified by contribution to total programme savings is appropriate. So that evaluation resources are used to provide greater accuracy regarding the largest portion of the tracking system’s estimated savings, the verification survey sample should be weighted toward programme participants thought to yield the largest percentage of total programme savings. This will result in verification of the largest percentage of estimated savings.

- Conduct follow-up telephone or mail surveys with participants, asking them to report:

    - The recommended measures they implemented, including the energy use of the measure (which may differ from the recorded level in the programme tracking system);

    - Other measures they may have implemented;

    - The degree to which measures were implemented as a result of the programme;

  - Details of all measure implementations that will permit a more accurate estimate of each measure's impacts;

  - Details about the participants that will allow extrapolation of the results to the entire tracking system of participants.

- Conduct follow-up verification visits with a sample of those participants who report having implemented specific measures (again, weighted toward those with the highest savings), to verify the type, number and level of the measure.

### 6.1.3 Follow-up Activities

The follow-up activities should yield two percentages:

- Reported actual savings compared to savings intended implemented – The percentage of the savings from measures participants said they planned or were likely to implement, as recorded in the tracking system, the percentage verified through the telephone/mail surveys.

- Savings verified on-site compared to reported actual savings – The percentage of the telephone-verified savings reported by each implementing participant that can be verified through the on-site visit.

These percentages can then be applied to the entire tracking system, to estimate the total energy impacts for the entire programme. The precision of this estimate, of course, will not be as high as for programmes that deal directly with measure implementation (e.g., incentive programmes) since the measures that were actually implemented must be estimated, rather than being recorded in the programme's tracking system.

The specific approaches to estimating impacts from each measure should follow the guidelines presented in Chapters 4 and 5, but it is likely that enhanced engineering estimation will be most suitable. Note that industrial firms are more likely to implement certain EE measures in order to obtain non-energy benefits (especially increased production or lower cost production), and may therefore be more likely to be free-riders.

### 6.1.4 Programmes One Step Removed from Implementation

Other targeted programmes, such as training programmes for trade allies, may be one step removed from measure implementation in that the implementation act itself occurs not by the individuals participating in the information programme but by consumers who are influenced by these individuals.

Well-designed training programmes will include brief surveys prior to and immediately after the training exposure, to document whether the participants' attitudes and intentions may have changed as a result of the training, and to record statements of the participants' intentions about behaviour leading to higher energy efficiency. There are two options for estimating the programme impacts:

- Contact the trade ally to identify influenced clients and contact these directly to assess the degree and character of measure implementation. This is more feasible with smaller information programme participant populations.

- Obtain estimates from the participating trade allies regarding the impact of the programme on the EE decisions of their customers/constituents. This information should be quantified and should include an estimate of the number of measures implemented, the efficiency level of those measures, and the actions that would have been taken in the absence of the programme.

With either approach, supporting data regarding the effects of the programme and of the trade allies should be gathered, if possible, including any other indicators that consumer/business decisions have changed. For example, if the participating trade allies are contractors, and the efficient measures they recommend to their customers are sold through retail outlets, the retails outlets can be surveyed to determine whether there has been an increase in sales of efficient units.

## 6.2 MARKET TRANSFORMATION PROGRAMMES

Market transformation is more an intended programme result rather than it is a programme type. The programme targets the removal or lowering of specific **market barriers{** XE "market barriers" **}** to higher efficiency. The appeal of market transformation programmes**{** XE "market transformation programme" **}** is that they attempt to cause *lasting* changes in specific markets, leading to higher efficiency purchases and behaviours. The underlying assumption is that since the targeted changes are lasting, additional programme intervention is not needed. This is in contrast to many traditional programmes, whose impacts on efficiency decisions last only as long as the programme intervention occurs.

### 6.2.1 TYPES OF MARKET TRANSFORMATION PROGRAMMES

Some example of programme types include:

- Manufacturer incentive programmes, in which incentives are used to reduce the risk and/or cost to manufacturers of developing or bringing to market more efficient products and technologies.

- Labelling programmes, which reduce the risk to consumers of knowing whether product efficiency claims can be trusted.

- Technology procurement programmes, in which manufacturers' risk in developing and bringing more efficient products to market is reduced through the use of a guaranteed or "highly likely" pool of purchasers who specify their exact product requirements (efficiency- and non-efficiency related).

- Mass market information programmes that seek to condition the market to accept and demand more efficient products and behaviours for a variety of market actors. Such programmes reduce risk to manufacturers and other trade allies who can provide the targeted efficient products and practices. They can also reduce transaction costs to consumers who face a need to better understand efficient products and their availability before they can feel comfortable purchasing them.

### 6.2.2 TWO MAIN COMPONENTS

Market transformation programmes have two main components that focus the programme evaluation:

- Direct energy savings – They often result in immediate energy savings stemming directly from programme activities, much as traditional EE programmes might. These must be quantified as part of the evaluation.

- Indirect energy savings – They intend to result in lasting market changes[14] which must be identified and tracked throughout the programme and for a period afterward, and the indirect energy savings they cause must be quantified.

To estimate **direct energy savings** from a market transformation programme, the evaluation must identify and quantify the number of participants and the efficient actions they may have taken. Once participants have been identified, the methods described in earlier chapters can be used to quantify the direct energy savings resulting from the programme. However, there is one difference: The level of resources allocated to estimating per-unit gross energy impacts should be minimised, so that evaluation resources can be focused on estimating market transformation impacts and attributing these impacts to the programme. As a market transformation analyst said, "The net effect of investing heavily in measurement of gross savings is therefore akin to estimating the total weight of a crowd of people by carefully weighing a sample of them, and then multiplying by a very rough estimate of the size of the crowd."[15]

The **indirect energy savings** from the programme – those resulting from the transformation of the market – are more challenging to estimate. The recommended evaluation approach is as follows:

- Ensure that the programme design has documented how the targeted market operates and how the programme is expected to interact with the market, including the key market actors, how they interact, the current level of efficiency, barriers, the type and nature of perceived market barriers, how the programme is expected to remove/lower them, etc. If it has not, try to reconstruct this information. This task is sometimes seen as an evaluation task, but it falls more logically into the role of the programme designers.

- Define the market indicators that will serve as evidence that the market is indeed changing. The market indicators selected should be logical and believable to the individuals who must make decisions about the programme.

- Estimate the baseline levels for these indicators e.g., through surveys with consumers and other key market actors such as product distributors and retailers, or possibly through an analysis of secondary data on market penetration or product sales. Develop thereafter a baseline estimate of the natural change in the market, in the absence of the programme. This should be based on an analysis of longer term trends in the market e.g., through an analysis

---

[14] *One might also think of this as permanent non-participant spill-over: Those not directly participating in the programme continue to make the efficient decisions promoted by the programme, because of the programme.*

[15] Evaluating Energy Efficiency Programmes in a Restructured Industry Environment: A Handbook for PUC Staff*, by J. Schlegel, M. Goldberg, J. Raab, R. Prahl, M. Keneipp and D. Violette, April 1997. See appendix for a summary of this document.*

of secondary data and/or interviews with key market actors such as manufacturers, distributors and retailers associated with the product.

- Conduct periodic surveys of the market (and use secondary data as appropriate) to measure the market indicator levels throughout the programme and shortly afterward.

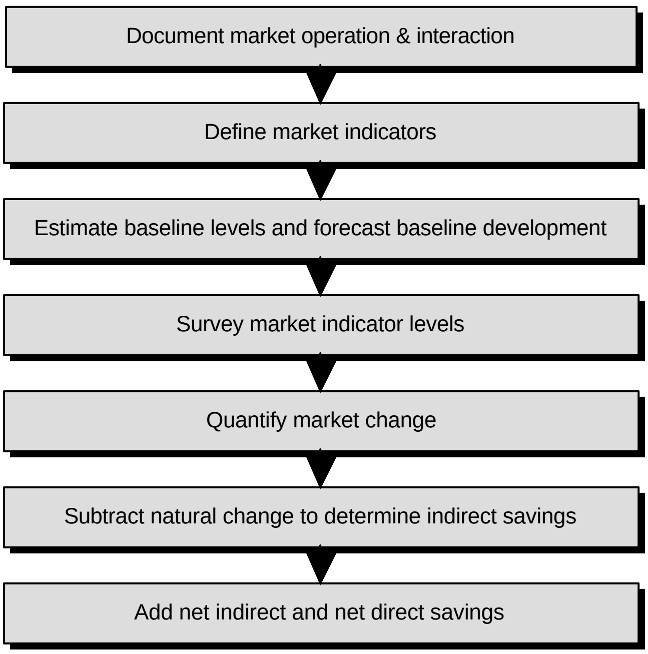
- Quantify the change in the market based on changes in market share or penetration and evidence of permanence in these changes.

- Convert the changes in market share/penetration to changes in efficiency for a specific number of products, so that the change can be expressed in terms of energy savings.

- Subtract out the effects of natural change in the market, to determine the net indirect energy savings resulting from the programme.

- Add these net indirect energy savings to the estimated net direct energy savings for the programme.

*Exhibit 6-1: Evaluation approach for indirect savings.*




**Forecasting the baseline development** is probably the most challenging aspect of the evaluation of indirect savings. One must forecast how a market will change in the future. Typically, the market is not at a steady state, and there may or may not be a steady trend that lends itself to easy extrapolation into the future. It is helpful to develop a number of independent estimates of this factor and make an informed judgement based on all available data. Certainly, product sales trends can play a part in this baseline market change forecast, as can reports prepared by various trade allies associated with the measure. What is not very likely is that the evaluation team will be able to develop a reliable natural change baseline for many years into the future (i.e., beyond the end of the programme). The use of a comparison/control country as a baseline is questionable, especially in the European context, since countries differ too much from each other. In any case, the evaluation team and the programme sponsor must agree upon decisions about the natural change baseline up front, with full admission of the uncertainties surrounding such estimation.

### 6.2.3 MARKET INDICATORS

By way of example, the following is a list of market indicators{ XE "market indicators" } defined for a programme seeking to transform the market for efficient lighting during lighting remodelling:

- Increased knowledge or awareness among planners, designers, and decision-makers about efficient lighting technologies (e.g. recognition rate).
- Existence and deployment of decision-making tools and structures which are likely to lead to efficient design and equipment installation, and which are being used on more jobs (number of businesses with energy management strategies).
- More frequent recommendation or specification of efficient equipment and design (e.g. publicity rate of efficient equipment).
- Increased sales/purchases of efficient equipment or design (e.g. sales statistics of efficient equipment).
- Increased application of efficient equipment or design (e.g. number of washes on economy programmes relative to normal washing machine programmes).
- Attendance at training and intent to implement training (e.g. number of training seminars and participants).
- Transfer of experience with efficient equipment and design to other buildings (e.g. inspiration source of EE initiative).
- Changes in the costs of efficient technologies and practices (e.g. retail prices).
- Changes in the equipment stocked by retailers/end-users.

The methods recommended for measuring these market indicators include interviews with vendors, contractors, and managers of targeted large firms, as well as walk-through surveys or plan reviews of samples of remodelled buildings.

### 6.2.4 "LIFE" OF SAVINGS

Some decision will have to be made regarding the methods for determining the number of years to assign to the "lifetime" of the market transformation effect. This, along with the above mentioned other elements of the evaluation, should be agreed upon up front with the key decision-makers.

Questions to be answered are for example:

- How permanent is a lasting change in the market?

- How does one account for new technologies being introduced into the market that may displace the one being promoted by the programme?
- Should the “life” of the market transformation impacts be considered as the number of years by which the natural diffusion of the technology into the market has been advanced by the programme?
- What is the basis for estimating that advance to the diffusion curve?

These factors must be decided and agreed upon. Conservative estimates are appropriate for high-profile programmes.

# 6.3 TRANSMISSION & DISTRIBUTION PROGRAMMES

As noted earlier, the basic concept of T&D**{** XE "T&D programmes" **}** DSM programmes is that, by delaying T&D upgrades through EE or load management programmes, the responsible utility can save substantial sums of money.

Evaluations of targeted T&D DSM programmes are relatively straightforward. However, as with most evaluations, their accuracy depends largely on the accuracy of the estimates of baseline conditions.

The utility’s supply and expansion forecast plans should include time-based estimates of customer load. For evaluations of T&D EE programmes, this forecast can be supplemented by market research, to collect data on the penetration of the targeted efficiency measures, intentions regarding purchases or behavioural changes promoted by the programme, and other free-ridership issues, to help refine estimates of baseline energy use. For evaluations of dispatchable load and load management programmes, supplementary data may be needed regarding intended use of the equipment or facilities targeted for curtailment during the programme’s dispatch period, to address similar concerns.

The mechanics of evaluating the programme follow those of evaluations of other EE or load management programmes (cf. Section 6.4). The difference for T&D programmes is that the “treatment” or “post-programme” estimate of demand and energy use can be based on readings at the substation level, using monitoring equipment installed at the substation and either recorded there or observed and recorded at a central facility. Substation measurement of energy use is not typically done for other types of DSM programmes, because effects at individual substations are usually too small to observe, amidst the normal variation in substation load. T&D DSM programmes are designed to produce significant demand reductions capable of being observed at the substation level.

# 6.4 LOAD MANAGEMENT PROGRAMMES

Load management programmes**{** XE "load management programmes" **}**, as a category of programmes distinct from other EE programmes, comprise those initiatives designed to reduce the instantaneous demand for energy, typically so that additional energy resources are not needed to meet the system’s energy demand. The objective is to avoid costly additional production and purchase of energy during **peak load periods** when energy costs are highest.

Examples of load management programmes are:

- Utility control of high-energy-intensity home appliances such as heating, air conditioning or water heating, or pumping equipment. Typically, the utility can send a signal to a control device, which either switches the targeted equipment off or cycles it.

- Utility control of similar equipment at commercial or industrial facilities.

- Thermal energy storage, in which equipment or systems use energy at off-peak hours and deliver it to the home or facility during peak hours, as needed.

- Timers on various types of high-energy-intensity equipment.

- Interruptible rate, curtailable rate, or standby generator programmes, which either turn off pre-specified energy end-uses at the customer's facility, request that the customer do so, or switch pre-specified equipment to generator-produced electricity rather than electricity from the power grid. In return, the customer is typically either paid an incentive or given a lower energy rate.

Key issues in evaluating load management programmes include:

- If equipment is turned off by the utility, for what percentage of the equipment does the "switching" work, i.e., what is the control system's reliability? It is unlikely that all switching systems will work perfectly, resulting in a portion of the "controlled" equipment not being controlled when the utility believes it is being controlled. The result is lower energy impact than expected or believed.

- Would the equipment being controlled during the peak period have been operating when the control strategy was activated, or would it already have been off (e.g., the homeowner would not have been home)? In such cases the programme would have had no real energy impact.

- If the equipment to be controlled would have been on when the control strategy was activated, at what level would it have been on, i.e., what would the actual load have been? While the programme should have recorded the rated capacity of the equipment, often equipment is not used at its full load capacity in a specific installation. The evaluation cannot assume that the equipment is operating at its full rated capacity.

Strategies for evaluating these issues include:

- For programmes requiring end-use/special meters, metered data can be analysed to determine operating conditions before, during and after the control period. Follow-up surveys can collect data on what load levels would have been like if the control period had not been in effect. The metered data can also be used to examine loads on similar days under similar conditions that can serve as a baseline for the control-period load.

- For load control programmes:

  - Use previous studies, if any exist, of the in situ load of the targeted end-use equipment, especially studies that included end-use metering to determine such loads. Per-unit energy impact estimates derived from such studies may also be of value as a cross-check on the evaluation's results.

  - Develop end-use simulations of prototype homes/facilities using and not using the equipment targeted for control.

- Meter a sample of the targeted end-uses before, during and after control periods. Ideally, the sample would include non-participant end-uses in addition to the participating ones.

- Conduct customer surveys to obtain data on:
  - Typical operating schedules for the targeted equipment/systems;
  - Customer characteristics that will allow data from matching non-participants to be analysed (e.g., for residential programmes - number of occupants at home during control period, appliance saturation; for commercial programmes - facility/business type, square metres, operating hours);
  - Whether controlled equipment was scheduled to be on during the control period;
  - Whether it was on just prior to the control period;
  - Whether it was controlled during the control period;
  - Operating settings and facility characteristics that would determine the percentage of rated capacity for the equipment (e.g., thermostat settings for space conditioning and water heating equipment, square metres of space affected by space conditioning equipment).

➢ In some circumstances, load data for participating and non-participating customers may be available (e.g., from cost-of-service load research studies):

- Load data from several days (e.g., 10) most similar to the control day(s) can be examined for both participants and non-participants.

- Weather data can also be collected and compared to similar data for a typical meteorological year (TMY). Load data can be adjusted to reflect the TMY.

- Participant usage can be subtracted from non-participant usage, to estimate programme impacts. Or, depending on the equipment controlled and the availability of data, participant usage during the control period can be analysed in terms of the similar uncontrolled periods to predict what usage would have been in the absence of the programme.

- Survey data can be used to confirm assumptions regarding operating schedules, facility characteristics, etc.

It is worth noting that the establishment of a baseline can prove very difficult in societies undergoing large changes in economy (see also Section 6.2.2). It is for example difficult at present to determine a reasonable baseline for the development in the energy sector in some of the East European countries. Evaluation of load management programme impacts compared to no intervention is in such situations problematic. The same holds true for new equipment markets in rapid development.

## 6.5 CUSTOMER RETENTION PROGRAMMES

The central objective of EE programmes operated for customer retention**{** XE "customer retention programmes" **}** purposes is **profitability**. In this sense they have the same goal as EE projects implemented by energy services companies. Energy impacts are secondary to the primary objective of retaining customers (or attracting new ones). In case evaluation of the energy impacts is carried out it is typically done to obtain public relations benefits. Evaluation of the energy impacts associated with individual projects is primarily done to satisfy the customer's need for assurance that the projected benefits actually are achieved.

The evaluation of central concern to the energy provider is of profitability: **Does the programme succeed in retaining customers and, if so, do the profit margins on the retained customers outweigh the costs of the programmes?** In competitive environments energy providers are likely to charge fees for many of their EE services, with the retention effect stemming from expanding the range of services available to the customer, the quality of those services and perceptions that the provider is looking out for the customer's best interests. The provider may discount certain types of services for certain customers who are deemed as having high value, and may provide some inexpensive services for free, as a benefit of being a customer of the provider.

In this context the evaluation of the customer retention programme is a study of overall profitability and, as a secondary objective, changes in market indicators of customer satisfaction. Customer retention programmes can be quite varied and so it is difficult to provide step-by-step guidelines for their evaluation. However, the primary issue to be addressed is that of whether the services provided (fee-based, free or discounted) function to retain customers who otherwise may have switched energy providers.

### 6.5.1 COMPETITIVE MARKETS

In competitive markets, evaluations should assess the many sources of profitability of long-term versus new customers (e.g., less expensive marketing costs, greater likelihood to purchase a wider range of services), to determine the extent to which retained customers provide such benefits for the energy provider.

**Retention rate benchmark studies{ XE "**retention rate benchmark studies**" }** can be performed i.e., the energy provider's retention rate before and after implementation of the programme can be compared to each other and to the benchmark.

In addition, the value of the specific customers retained through each year of the programme can be analysed on the basis of their contribution to profit margins and the cost to serve them.

### 6.5.2 MARKETS IN TRANSITION

For markets that are making a transition to retail competition, the analysis is somewhat speculative. How does one measure the success of a programme to retain customers when customers do not yet have a choice of energy supplier? Most such studies rely on:

- Indicators of customer satisfaction and likelihood of switching suppliers for various discounts in energy price or various combinations of price discounts and other energy services (using survey research).
- The change in the company's profitability resulting from customers who are persuaded to sign long-term contracts, often at discounted energy prices (i.e., profits relative to having the customer defect to another energy provider).

The problem with the first of these indicators is that one never knows what the competition will be offering and how the customer will really respond to it. Satisfaction levels can be very high and customers can report that they are not at all likely to switch providers under a number of specified assumptions. However, these assumptions may not hold. The customer may be offered a product or service, or combination of products and services not anticipated in the research.

The customer's preconceptions about the nature of firms that might offer the competing services may be different from what actually would occur (e.g., the customer may be envisioning a neighbouring utility company when the real competitor is a highly regarded national provider of a different type of product who expands into the energy services market).

The customer may express some level of dissatisfaction with the current energy provider and even a willingness to switch under the circumstances posited during a survey, **but fear of the unknown and a low tolerance for risk** may prevent the customer from following through with such an action.

On the other hand, customers could be tempted to switch just because it becomes a possibility which is given massive media attention and other customers appear to switch.

Results of such studies are tentative at best, and best used for identifying areas for improvement rather than quantifying the profitability benefits from the customer retention programme.

Other issues to be addressed include:

- How any fees charged for EE services compare to the costs of providing those services.
- Whether the pricing points for various services maximise profitability.

## 6.6 ENERGY SERVICE COMPANY PROJECTS

In relation to EE projects performed by energy service companies**{** XE "ESCO projects" **}** (ESCOs), evaluation serves a monitoring and verification (M&V) function. The nature, methods and costs of this M&V effort are defined in the contract between the ESCO and that provider's customer, with the M&V activity serving as a basis for payments from the customer to the ESCO.

There are three interlocking **objectives** of M&V efforts:

- Determine the amount of the customer's payment to the ESCO.
- Provide information that will aid in operating the facility more efficiently.
- Assess whether the EE measure is performing as expected.

Ideally, an outside agent, independent of both the ESCO and the customer, should perform the M&V. However, in most cases the ESCO provides this function, to simplify execution and control costs. The extent of the evaluation depends on the level of uncertainty and perceived risk on the part of the customer; the higher the perception of risk, the more important and more extensive M&V is likely to be. M&V becomes a method for allocating risk between the ESCO and the customer.

The customer has a vested interest in spending as little as possible to reap the financial rewards of the EE project and knows that the M&V represents a cost that must be subtracted from the financial benefits it will receive. Chapter 4 explores the various conditions, which cause more extensive and comprehensive M&V.

The **three typical steps** in the M&V process are as follows:

- Verify the baseline.
- Verify the installation and correct operation of the EE measure.
- Verify the continued operation of the measure at regular intervals.

For well-understood measures being implemented at customer facilities having sophisticated facility managers or even energy managers, some or all of these steps may be bypassed.

A guidebook published in 2000 by the U.S. Department of Energy suggests that there are four options for M&V of ESCO projects[16]. These are presented in Exhibit 6-2 below.

*Exhibit 6-2: M&V Options*

| Measurement & Verification Option | How Savings Are Calculated | Cost |
|---|---|---|
| **Option A**: Focuses on physical assessment of equipment changes to ensure the installation is to specification. Key performance factors (e.g., lighting wattage or chiller efficiency) are determined with spot or short-term measurements and operational factors (e.g., lighting operating hours or cooling ton-hours) are stipulated based on analysis of historical data or spot/short-term measurements. Performance and proper operation are measured or checked annually. | Engineering calculations using spot or short-term measurements, computer simulations, and/or historical data. | Dependent on no. of measurement points. Approx. 1-5% of project construction cost. |
| **Option B**: Savings are determined after project completion by short-term or continuous measurements taken throughout the term of the contract at the device or system level. Both performance and operations factors are monitored. | Engineering calculations using metered data. | Dependent on no. and type of systems measured and term of analysis/ metering. Typically 3-10% of project construction cost. |
| **Option C**: After project completion, savings are determined at the "whole-building" or facility level using current year and historical utility meter or sub-meter data. | Analysis of utility meter (or sub-meter) data using techniques from simple comparison to multivariate (hourly or monthly) regression analysis. | Dependent on no. and complexity of parameters in analysis. Typically 1-10% of project construction cost. |
| **Option D**: Savings are determined through simulation of facility components and/or the whole facility. | Calibrated energy simulation/modelling; calibrated with hourly or monthly utility billing data and/or end-use metering. | Dependent on no. and complexity of systems evaluated. Typically 3-10% of project construction cost. |

Sophisticated purchasers and those with significant capital resources have little reason to use an ESCO since the ESCO adds little value.

Unsophisticated, capital-constrained purchasers have more interest in arrangements that can produce an energy cost savings stream derived from a performance contract, such as ESCOs offer. These customers may pay more for their EE measures but many ESCOs can offer contracts that (1) keep the cost of the measure off the balance sheet of the customer and (2) shift

[16] International Performance Measurement and Verification Protocol, *U.S. Department of Energy, 2000. This document is available on the World Wide Web at http://www.ipmvp.org*

the risk of the efficiency measure not producing the expected savings to the ESCO{ XE "ESCO" }. It is where project risk is shifted to the ESCO that M&V is most important. The ESCO must prove that savings have accrued in order to be paid, and the ESCO consequently charges the customer for that proof.

The methodologies described in the International Performance Measurement and Verification Protocol are certainly state of the art for monitoring and verification of the savings of ESCO{ XE "ESCO" } projects, and therefore also applicable in the EU. They are, however, very general, and require project-specific adaptation. In many EU Member States, technical rules, norms, or guidelines for assessing the energy consumption of buildings, heating, ventilation, air conditioning, lighting, production plants, etc., do exist, but may not be known to potential customers of energy performance contracting projects and third party financing projects.

A compilation of a national common set of existing or new technical rules, norms, or guidelines for assessing the energy consumption of buildings, heating, ventilation, air conditioning, lighting, production plants, etc. will be able to function as a national "reference guideline" to both ESCOs and potential customers for monitoring and verification of the savings of ESCO projects.

**Case Example: DSM Bidding Pilot Programme**

After the introduction of a competitive market, Stadtwerke Düsseldorf offered a DSM bidding pilot programme to medium-sized industrial and commercial customers. An example of evaluation strategy for a project granted support is given in the following.

The bidder (ESCO) who won the contract was the building management unit of the client. The client was a large service sector company.

Energy conservation measures:

- Reduction of air leakage by closing "short-cuts" between air inflow and air outflow;
- Closing down 7 fans that are no longer needed after the reduction of the leakage;
- Installation of variable speed drives in the remaining 12 ventilation fan motors to reduce the circulating air quantities as well as the electricity demand further.

The bidding company implemented the measures itself, so no third party financing took place. Hence, the bidder had to verify the savings to Stadtwerke Düsseldorf to get the full award payment. Wuppertal Institute proposed the methods for verification of the savings based on the IPMVP. More precisely a mix of Option A and Option B was proposed:

- For the measurement of the situation before measures, Option A was chosen. It was proposed to measure for a short term the actual value of the power input and the air volume of a representative of each of the three types of fan/motor systems that were present among the 19 fans in total. This was justified because the motors were the same type and size and running continuously before the refurbishment.
- For the measurement of the situation after measures, Option B was proposed and chosen by the bidding company. The original proposal was to make short-term measurements of the power input and the air volume, and to monitor the operating hours of each of the fans over a longer period. This was needed because the 7 fans, which were closed down, still remained in place as back-up for defects or exceptional heat loads. However, the company found an even cheaper and better way to monitor the energy consumption: It simply installed 2 meters into the 2 electric circuits that exclusively feed the 19 fan/motor systems, and continuously measured the consumption using the building automation system in place.

**Wuppertal Institute, Germany**

# 7 PROCESS AND MARKET IMPACT EVALUATION

Much has been written about the survey techniques used in process evaluations and market impact evaluations, because they are in large part identical to those used in **market research**. Their use in the evaluation of EE programmes is different primarily with regard to the content of the research questions addressed, but not in how the methods are implemented or how the results are analysed. This chapter briefly describes the techniques used[17]. For more information on survey techniques**{** XE "survey techniques" **}** the reader is referred to the evaluation research references documented in Appendix B of this guidebook.

Surveys related to process and market impact evaluation may for example address:

- Efficiency of programme procedures, outreach, and information processing.
- Methods for streamlining the programme and improving cost-effectiveness.
- Explanations behind programme impact estimates (i.e., how and why the impacts were as estimated).
- Market segments that participate and do not participate in the programme.
- Effects of the programme on equipment manufacturers, suppliers and market channels.
- Participant satisfaction with the programme.
- Effectiveness of marketing strategies and promotional materials.

## 7.1 PRIMARY SURVEY TECHNIQUES

The primary survey techniques**{** XE "primary survey techniques" **}** used are as follows:

- In-person interviews — These are often conducted with programme staff and sometimes trade allies or implementing contractors, to provide a comprehensive understanding of how the programme is actually operating.
- Analysis of the programme tracking system — This is used to better understand which market segments are actually participating in the programme and to develop sample frames for participant and non-participant/comparison group surveys (see also Section 4.1.1).
- Telephone, mail or on-site surveys with customers — These often serve both demand impact estimation and process/market evaluation purposes, and are used to gain information on how the programme is operating, from the customer's perspective.

[17] *Part of the following text is inspired by "The Swedish Evaluation Guidebook" (NUTEK et. al., 1993). The structure has, however, been modified and the text concentrated and re-arranged with some small additions/changes.*

- Secondary market research — This method is sometimes used to collect information on how markets operate, and to construct baselines for market indicators that may be useful in estimating the market impacts of programmes.

- Other techniques sometimes used in the context of EE programme evaluation include:

  - Mystery shopper analysis, in which researchers pose as potential purchasers, to determine how sales people portray EE products or the concept of energy efficiency to potential purchasers.

  - Research panels, similar to focus groups, except that the number of individuals attending the group can be much larger (e.g., 30-40) and the group may include both quantitative (e.g., written questionnaire) and qualitative (e.g., focus group-like discussions with smaller subgroups) research.

## 7.2 DIRECT OBSERVATION AND MEASUREMENT

Direct observation**{** XE "direct observation" **}** and measurement has the advantage of not relying on respondents' ability and willingness to answer. To the degree that it is possible to observe/measure that which you want to know, the answers you receive will be of much higher quality than from questions put directly to the respondents.

The disadvantage of direct observation is that you have no way of finding out what the target group likes or is thinking. In addition, the method can be rather expensive. For example, it is possible to observe that energy consumption has changed, but finding out why this has occurred, probably requires asking the consumers. Therefore, direct observations are often combined with surveys.

Some examples of which questions regarding the "energy market" the method could be used to answer are:

- Who pays their bills at the customer service office?

- Who visits the customer service office?

- Who participates in voluntary activities (like energy clubs)?

- How much energy does the household consume?

- How do time of day tariffs affect energy consumption?

- How does energy consumption in an area vary due to the programme?

It is often advisable to combine the results of direct observation with surveys, but the use of the direct observation is often decided by the ease of performing it and the advantages it has over using only surveys.

# 7.3 PERSONAL INTERVIEW, TELEPHONE SURVEY, OR MAIL SURVEY

A survey is typically a **randomly selected investigation** with the purpose of describing a target population. A survey is carried out when a representative picture is desired of what the market thinks, feels, or does.

The data collection methods chosen depend on the existing circumstances. How many resources are available? How many people do you need to reach? Do the questions require explanations or modifications for each individual respondent? Do you want an immediate reaction, or is it more important that the respondents have time to think over their replies? How much time is available before the survey results must be presented?

Below advantages and disadvantages are listed for each of the three methods personal interview**{** XE "personal interview" **}**, telephone survey**{** XE "telephone survey" **}**, and mail survey**{** XE "mail survey" **}**.

## 7.3.1 PERSONAL INTERVIEWS – FLEXIBLE BUT EXPENSIVE

### *ADVANTAGES*

- The interviewer and the person providing the information meet face-to-face, which is the most flexible situation.
- The interviewer has an opportunity to explain things.
- Many questions may be put forward in a short time.
- Virtually all kinds of questions may be asked - even sensitive ones - if special measures have been taken (such as an anonymous, written section included as part of the interview).
- Open questions may be asked if the interviewer is skilled enough to follow up on the replies.

### *DISADVANTAGES*

- Personal interviews are very expensive.
- It may be difficult to limit the responses to the intended respondent (that is, other persons may be present).
- There is much room for non-verbal communication that is difficult to control.

## 7.3.2 TELEPHONE SURVEYS – SIMPLE RESPONSE CATEGORIES BUT QUICK

### *ADVANTAGES*

- Less of a problem concerning the presence of others.
- Only verbal communication takes place but pauses can be troublesome.

- Less room for bias from the interviewer.
- Most topics can be addressed with the possible exception of the most sensitive.
- A large number of respondents can be contacted in a short amount of time.

***DISADVANTAGES***

- More limited social interaction, the interviewer remains anonymous. Less flexible.
- Only a limited number of questions can be asked.
- Open-ended questions are more difficult, and the interviewer has limited time for writing them down.
- Simple response categories are required, since the respondent can't look at the available choices.

**Case Example: Evaluation of the Energy Efficiency Check**

Our initial strategy was a qualitative survey using 400 telephone interviews including a control group. However, based on the Guidebook and expert advice we realised that it would be difficult or impossible to establish a control selection or "baseline". Based on the vintage of the programme and an assessment of, which questions we wanted to answer, we decided to perform a quantitative survey with 1,200 telephone interview to ensure significant results. We chose a main sample and two "control" samples to answer our questions regarding effects related to "non-participants", self selection, rebound, free-riders and spill-over effects, etc. As a result we decided to use an agency specialised in market analyses and not the initially chosen agency specialised in psychology and depth interviews.

**Norsk Enøk og Energi AS, Norway**

## 7.3.3 MAIL SURVEYS – INEXPENSIVE BUT LOW RESPONSE RATES

***ADVANTAGES***

- No interviewer bias, but on the other hand, no interviewer who can explain.
- All topics are possible, even sensitive ones.
- Inexpensive way to gather data, especially when a large number of respondents are desired.
- A relatively large number of questions may be posed.

***DISADVANTAGES***

- The initial contact is made exclusively via a cover letter and the questionnaire itself (least flexible situation).
- Difficult to assist the respondent (such as with difficult questions).
- Open-ended questions are hard to use.

- The response rate can be low.
- A long time is needed for collecting the answers and a large number of follow-up mailings are typically required.

**Case Example: National Energy Efficiency Programme**

The Czech Ministry of Industry and Trade has a national financial support programme to encourage EE. Subjects applying for support are obliged to carry out an energy audit, describing the actual state of energy consumption and the energy saving possibilities. Demonstration projects approved by the CEA receive 40% support while replicated projects receive 15% of investment costs.

An annual monitoring report specifying the energy consumption by type of energy for the past year must be submitted to the Czech Energy Agency. A list of simple questions guide the recipients in making their monitoring report. Part of the contents of the monitoring report is based on the recipient's energy bill (which in the Czech Republic contains information on energy units consumed, unit price, and total cost per energy type).

One of the lessons learned from the ex-post evaluation was that the correctness of the values of the indicators in the monitoring reports was questionable:

- Some mistakes were caused by incorrect conversion of energy units (e.g. $m^3$ of natural gas conversion to GJ). The monitoring requested conversion to GJ, which appeared not to be a straightforward task.
- There was irregular application of weather adjustment in the monitoring reports.
- The price of fuel was not filled in correctly taking into account the different prices of fuel in the individual districts and the price changes over time.

The evaluation clearly proved a need for improvement of the monitoring system. The guiding questions for the monitoring must be further simplified and reassessed. The monitoring reports are not prepared by professionals (contrary to the tender documents). The indicators to be included in the annual monitoring report should therefore be simple, so that the project responsible does not have to carry out the slightest recalculations or adjustments.

**SEVEn, Czech Republic**

## 7.4 IN-DEPTH AND GROUP INTERVIEWS

Another group of market surveys is based on exploratory methods. An exploratory survey is carried out whenever there is not enough knowledge about the survey topic. These are most often qualitative surveys using in-depth**{** XE "in-depth interviews" **}** and group interviews**{** XE "group interviews" **}**. Their purpose is to find out why people act in certain ways.

This type of survey is thus focused on individuals and their way of reacting, which is different from descriptive surveys, which are question-oriented. Personal interviews present great opportunity for exploring in-depth **why** an individual reacts in a certain way. The method puts a lot of responsibility on the interviewer to interpret the replies. Conversely, the aim of a question-oriented survey is to find out what proportion of individuals (i.e., **how many**) have reacted in a certain way.

The most common type of qualitative study is the **in-depth interview**. In this case, the respondent (the person being questioned) has a lot of freedom in addressing the relevant topics, while in market surveys the interviewer retains the initiative by directing the respondent's attention. The interviewer often has a list of specific points to be taken up, but it is important to leave room for in-depth follow-up questions that can lead to further insight and understanding.

It is also possible to do the same kind of interview with a group of respondents under the guidance of a question leader, that is a **group interview**. There is an abundance of group interview techniques that may be used. The survey leader responsible for the interview should be trained in carrying out group interviews. It is important that the answers are not steered, while at the same time the discussion should not become side-tracked from the intended topics. Since group members influence each other, experience has shown that the best results come from building groups that are relatively homogenous.

A common view is that qualitative surveys should be a complement to other surveys in order to provide better insight into marketing problems. In-depth and group interviews should therefore be used along with other surveys, business evaluations, and creative inspiration. The goal is not only confirming or discarding results that have been obtained, but also to open new angles of attack.

Since the number of interviews is low (inexpensive evaluation) and statistical representation of the sample is not required, these methods are especially suitable for preliminary studies. However, qualitative surveys should not replace quantitative surveys, even if it is tempting to hope that the results are sufficiently representative of the entire target group.

An important issue is the interpretation of the answers. It is easy to interpret the answers so that they correspond to what you want to hear, which is why every attempt should be made to use a group leader trained for the task instead of trying to conduct a group interview by yourself. In addition, the less steered the interview, the harder it is to compare the various respondents' replies.

# 7.5 QUESTIONNAIRE DESIGN

Questionnaire design**{** XE "questionnaire design" **}** consists of three steps as illustrated by Exhibit 7-1.

*Exhibit 7-1: Questionnaire design process.*


## 7.5.1 PROBLEM ANALYSIS

The point of departure for all market surveys is problem analysis. It is extremely important to do a proper problem analysis before formulating the questions. The analysis should clearly state, which issues may influence the topic to be surveyed.

Reading through a well-executed problem analysis, it is easy to understand why all items of the resulting survey instrument were included. In addition, it should be clear why certain other items were not included, items that may appear to be important in the given context.

**Carefully think through the problems to be surveyed. Specify concepts and get an overview over which variables are essential for the survey before you begin to formulate the actual questions.** Too often people start writing concrete questions too early.

The definition of the problem area, question structure, sampling considerations and data collection methods are all related and affect each other. The questions of a questionnaire can be formulated in many different ways depending on which segment of the population is to be questioned and which method for gathering the data is to be used.

It is difficult to make general rules about how this important, initial work is to be carried out. Creativity, imagination, and the ability to put oneself in someone else's place are important ingredients during the planning stage. Also, it is not really possible to give rules of thumb on how one should go about finding the best approach for a survey or working out the best design or strategy. The best training for this kind of work is, presumably, to read and study how these problems have been solved in previous surveys.

Furthermore, it is wise to decide whether the survey will be carried out as face-to-face interviews, telephone surveys or mail surveys, before writing the questions.

**Case Example: Improving the Heating System Balance in Buildings**

The objective of the evaluation was to estimate the heating system situation after programme implementation.

The first step of the evaluation was collection of information on programme and the follow-up studies and interview of partners involved in the programme. Based on this information, two questionnaires were developed targeted at the house managers involved in the programme (second step). The objective of one questionnaire was to collect information on the building and the programme related renovation of the heating system while the objective of the other questionnaire was to collect information regarding participant satisfaction and energy awareness level.

And finally the last step was collection of energy consumption data for the individual buildings from the district heating companies.

**Motiva, Finland**

### 7.5.2 DESIGN OF QUESTIONNAIRES WITH MULTIPLE-CHOICE ANSWERS

The instrument applied in a market survey of energy services is commonly a standardised questionnaire. The main advantage of a standardised questionnaire, i.e., a questionnaire with multiple-choice answers**{** XE "multiple-choice answers" **}**, is that these are easier to encode and analyse than questionnaires using many open-ended questions.

In the actual construction of the questionnaire, the content, the structure, and the order of the questions; the layout; the respondent characteristics; and review of the topics once more have significance on the replies. Below, a checklist for the design is given.

**Question content, structure, and order:**

- Consider a mix of data collection methods from the very beginning of the design process.

- Reflect on the order of the questions. The basic recommendation is to begin with broadly worded questions and proceed to more detailed ones. There should also be a logical sequence to the sub-topics. In other words, a “train of thought” running through the questionnaire. Especially sensitive questions should be saved to the last so that the respondent is not frightened off.

- Think about the possible effects of preceding questions and replies and thus the right question order. For example, once you have made the respondent aware of a price, this will bias the replies to the remaining questions.

- Consider the amount of effort the respondent will need in answering the question.

- Use open-ended questions only when really needed.

**Layout:**

- Take advantage of the opportunity that interviews and mail surveys present for using illustrations to make the content of quantitative reply alternatives more clear.

**Formulation:**

- Don’t include questions just because they might be nic e to have.

- Define the questions in place and time.

- Ask only about one thing at a time.

- Be careful when using “yes/no” questions. In most cases more alternatives are needed.

- If predetermined multiple -choice responses are used, then the available choices should be exhaustive, mutually exclusive, and easily comprehensible.

- Avoid unbalanced questions that gives a certain answer a positive/negative bias; or alternatively, use several questions with opposite emphasis.

- Ask secondary questions that touch on the thoughts and notions that lay behind the attitudes of the answer to primary questions.

- Be careful with hypothetical and retrospective questions. The replies may be very difficult to interpret.

- Avoid value-laden words and leading questions.

- Be as concrete as possible.

**Respondent characteristics:**

- Is the respondent able to reply to the question? When asking about one’s knowledge of something that only some of the respondents will be able to reply to, there should be a “don’t know” alternative to multiple choice items.

- Is the respondent able to express his thoughts and opinions in words? Don’t force an opinion from the respondent unless there are specially motivated reasons for doing so.

- Will the respondent actually do what he says he will? What the respondents say they are going to do often differs from what they actually do. The relationship between the respondent's plans and actions will be greater the more important the issue is to the respondent. Thus, questions concerning the respondent's intentions should only be asked when important decisions are concerned.

**Review the topics once more:**

- Have you considered how the results are to be used or how the results are to be analysed?
- Will the questions really answer what you want to know?
- Are the questions so well thought out that they can be used in the future for comparing results between different years? The slightest little revision in the wording can result in making it impossible to compare the replies to previous years.

### 7.5.3 PRE-TEST THE QUESTIONNAIRE

**Always pre-test a questionnaire!** All interview questions must be tested. A survey's efficiency is dramatically improved if the draft questionnaire undergoes various "desk top tests" and pre-testing in the immediate environment even before testing it in the field.

This also applies to borrowed questions if they are being used in a new context. It is important to use every available opportunity to pre-test the questionnaire in the field. The efficiency of the questionnaire is further improved if you have a clear idea about just what it is you wish to test. You should also seize the opportunity to work on reducing the size of **measurement errors**.

There are many reasons why measurement errors are made in market surveys. Measurement errors may be classified as **interviewer bias{** XE "interviewer bias" **}**, instrument effects, or respondent effects.

Measurement errors stemming from interviewer bias show up when different interviewers get different answers or replies when asking the same question to the same individual. Studies have demonstrated that one interviewer may systematically obtain more positive replies than other interviewers. Some causes for interviewer bias include:

- The interviewer communicates ideas to the respondent that biases how the respondent replies.
- The interviewer's outlook is such that he interprets the replies in a different way than others would interpret the same replies.
- Emphasis given to different words can bias the replies.

Measurement errors stemming from **instrument effects{** XE "instrument effects" **}** are common in marketing connections. Some causes of instrument effects include:

- "Double-barrelled" questions that measure several underlying variables at the same time, while the person using the question believes that it only measures a single variable.
- Ambiguous wording (Is "listening to the radio" the same as "having the radio on"?)

- Vague wording (such as "regularly" and "sometimes") will mean different things to different people.

Some causes of **respondent effects{** XE "respondent effects" **}** include:

- Inability to answer (doesn't know, forgot, cannot state reasons for inability).
- Unwillingness to answer (invasion of privacy, lack of time, fatigue).
- Reluctance to answer truthfully (prestige seeking and thus give socially acceptable replies, a desire to be polite and co-operative, etc.)

All of the measurement errors named above obscure the survey results. Things that may appear to represent very small differences in wording can lead to completely different results; for example as in the difference between "should forbid" and "should not allow." Ignorance of how these sources of error can bias the results translates into inability to correctly interpret the results. Finally, it must be recognised that one can never completely eliminate all sources of error, and thus the goal should be to eliminate as many as possible, specially the major ones.

**Case Example: Evaluation of the Energy Efficiency Check**

We found that we had used a "leading question" in our investigation, which appears to have given misleading results. The question was "Do you *remember receiving* an EE newsletter within the last 12 months?" Here 47% of Group 3 answered positively although they had not been sent the EE newsletter or other EE material from the EE Centre or others in the area within this period. The question was leading in that it gave the impression that the newsletter had been sent and the question therefore only concerned whether or not the interviewee could remember or not.

The order of questions is also incredibly important. An investigation can be ruined completely if "revealing" questions are asked too soon.

It is important that persons who know the EEC programme and the use of it are involved in the design of the questionnaire. Many phrases were changed and many questions corrected in the questionnaire used by the market analysis bureau. This has increased the total quality of the investigation.

**Norsk Enøk og Energi AS, Norway**

The pilot testing also creates the opportunity to fine tune the questionnaire wording and try out different phrases and sentence constructions. The checklist below points out several ways to **work with the language{** XE "language" **}**:

- Define difficult words and use sophisticated words carefully.
- Edit away unnecessary words and parts of words.
- Don't be inflicted with "noun-itis."
- Use personal pronouns and other short words.
- Avoid abbreviations.
- Mix long and short sentences.
- Write the same way as you talk. Read the questions out loud to yourself and others.

- Is the text to be read or heard?
- Try to measure the text's readability.

It is always the author's fault when the reader has difficulty understanding a text.

## 7.6 RESPONSE RATES

**One should never expect a response rate{** XE "response rate" **} of 100%** when carrying out a survey. There are many reasons for lack of response. Prior to calculating the response rate in some surveys cases, commercial survey firms like SIFO[18] will exclude a large number of categories from the survey population, including the deceased, those residing abroad, the incarcerated, the hospitalised, non-speakers of the native language, etc. In addition to this, the percentage not responding to personal interviews is typically 25% due to no contact being made or refusals.

The response rate will be affected by the subject matter of the questions and the questionnaire's appearance, including the number of items. If the respondents feel that the subject matter is stimulating and relevant, then not only will the response rate be higher, but it will be possible to include a greater number of items. Another significant factor is how much the respondent trusts the party asking the questions.

**A reasonable goal for evaluations of energy services is a response rate of at least 50%.** Most households appear to be very interested in their energy costs, which means that a high response rate may be expected. You should also find out whether the response rate is the same among all groups. For example, if the non-response rate is especially high in the "major consumer" group, extra resources should be committed to acquiring the responses from this group. Unfortunately, there are often high non-response rates for certain subgroups in many surveys, which distorts the results (self selection bias, see Section 4.2.4 for more information).

## 7.7 INTERPRETING THE RESULTS OF INTERVIEWS AND SURVEYS

One of the most challenging tasks in a process or market evaluation is interpreting the results of the research. Those being interviewed or surveyed at times may try to answer questions in ways that they believe will:

- Please the interviewer.
- Make themselves seem intelligent.
- Make it seem that others are responsible for problems they may have created.

Research respondents may also often not know or remember the answer to questions the research must find answers to. The classic example of this problem is evaluators asking programme participants one year after they implemented a measure promoted by the programme what they would have done if there had not been a programme (the classic free-

[18] Svenska Institut för Opinionsundersökningar *(The Swedish Institute of Public Opinion Research).*

ridership question). Often, the respondents do not consider the measure implementation to be a defining moment of their lives. They may or may not remember what they were thinking one year ago. Even if their participation decision is more recent, it may be difficult for them to accurately describe what they would have done if there had been no programme.

Process and market evaluators address issues that may arise due to the **self interest{** XE "self interest" **}** of programme implementers by asking multiple parties involved in programme implementation similar questions and then thoughtfully weighing the responses they receive from each respondent type. They also address most of the above issues through careful questionnaire design.

Sources of information on how to address other process and market evaluation issues include the proceedings of the major EE conferences and the primary reference sources noted in Appendix B of this document.

# 8 APPLYING RESULTS

An important issue in relation to EE activities is the documentation of the process and the results together with the assumptions. The EE activities are often subject to public scrutiny and it is therefore wise to clearly document every activity including assumptions and choices to avoid unjust criticism. The same holds true for the associated evaluations.

## 8.1 PRESENTING RESULTS

The aim of presenting evaluation results may include one or more of the following:

- Proof/verification of estimated impact and costs;
- Basis for improved programme efficiency;
- Prepare grounds for new EE activities;
- Permission to continue programme efforts or initiate new programmes;
- Release payment for services rendered;
- Encouragement of new finance;
- Transfer know-how;
- Encourage the interest in energy efficiency EE (e.g., amongst non-participants).

A major issue in evaluations – and especially process evaluations – is how to present results. Programme staff, who are the most likely to be able to use the results of this research, may also be the most threatened by it, for it may expose ways in which the programme can or should be improved or terminated. For some programme implementers this may be **unwanted criticism**. For this reason, results need to be presented in a balanced manner, describing both the ways in which the programme is being operated successfully and the ways in which it can be improved. It is also helpful to provide immediate **feedback** on the programme, with regard to areas requiring improvement, as the process evaluation**{** XE "process evaluation" **}** proceeds. This offers the implementers an opportunity to make immediate improvements, which then can also be reported in the process evaluation results. In this way the process evaluation can be used as a tool of programme management.

When used strictly as an assessment of how well individuals are performing their jobs, the evaluation is likely to be perceived very negatively, limiting the willingness of programme implementers to share information, especially information about ways in which the programme is not living up to expectations.

For all types of evaluation, it might prove beneficial to arrange a review of the evaluation plan by several **interest groups** prior to the actual evaluation. This approach is likely to increase the general acceptance of the programme and the programme results. Furthermore (and this may be just as important) it may identify weak elements in the planned evaluation and allow timely improvement of the evaluation plan.

Target groups for dissemination of selected elements of evaluation results or all evaluation results include:

- Programme designers and implementers;
- Programme partners and allies;
- Political key-decision-makers;
- Finance providers;
- Participants and non-participants;
- General public;
- Other EE experts.

The target groups of the results will determine how and when results should be presented.

## 8.2 TIMING OF RESULTS

The usefulness of the results of many evaluations conducted in the past has been hindered due to poor timing{ XE "timing" }. Thorough research has been conducted, sometimes resulting in important discoveries about ways in which programmes should be refined, but the results have been presented after key decisions have been made about the next year of the programme. This has meant that programme changes based on the evaluation of the first year of a programme sometimes could not be made until the third year of the programme. This has devalued the evaluation to decision-makers, for they may have paid for a comprehensive evaluation and received valuable feedback on the programme too late for that feedback to be of use. More on this issue can be found in Section 3.7.

One way in which the cost of evaluations has been minimised and their usefulness maximised has been to tie specific evaluation activities to performance indicators. All major programmes should receive at least one comprehensive evaluation. However, for programmes lasting some years, evaluations can be linked to a set of indicators regarding programme performance, so that limited and very targeted research is conducted continuously but on an ad hoc basis when needed.

**Performance indicators**{ XE "performance indicators" } are typically quantifiable indices of how well a programme is performing, e.g., average expenditure per participant, number of participants per quarterly period, average tracking system estimated energy savings per participant, or average time between participation application and EE measure installation. An expected value is established for each performance indicator. If the programme does not meet this **threshold value** during the measurement period (e.g., quarterly), some form of evaluation research is triggered, such as telephone surveys with an appropriate sample of the participants. The research focuses specifically on the issue represented by the performance indicator, and its purpose is to determine why the indicator is not reaching its minimal threshold value.

Rather than wait until the end of a programme year to conduct the research, it is conducted – only if necessary – throughout the year, in short spurts. Programme implementation staff therefore receive rapid feedback so that they can alter programme components in time to affect programme performance in the same year. Performance indicators are thus a form of programme monitoring points designed to enhance programme performance *in a timely manner*, to focus the efforts of programme implementers, to minimise the cost of evaluations, and to maximise their usefulness to decision-makers.

## 8.3 TRANSFER OF PROGRAMMES

International exchange of experience can function as good inspiration for new and improved EE programmes. However, programmes cannot be transferred{ XE "transfer of programme" } directly from one context to another and be expected to result in the same outcome.

The local context is determined by the characteristics of the energy market, the energy utility industry structure and ownership, the EE market and EE providers, and the regulatory framework. The history is also relevant to consider.

A sound understanding of the programme mechanisms and problem mechanisms (see Section 2.5.1) will help identify the characteristics of the energy end-use, which the programme aims to influence. Often the experienced EE programme expert knows to some extent intuitively what will work in his/her local context and what not. Sometimes he/she will even have a good understanding of why. But often this knowledge is not included in the reports on the programme design, implementation, and outcome. Therefore, it is very useful to contact the person responsible for the programme, which you consider copying, and discuss the details of the underlying basic assumptions – in particular about consumer and market motivations.

Are the investments in a new programme going to be significant, it might prove wise to test some of the basic assumptions, for example, by implementing a pilot programme first.

# INDEX