1 PROCESS AND IMPACT EVALUATION

Process evaluation = Systematic assessment of the programme for the purpose of improving its design, its delivery, and the usefulness of the quality of services delivered to the consumer.

Impact evaluation = Evaluation of the effect/outcome (changes of behaviour and energy saved).

Please note that different classifications exist but the taxonomy is not important in itself. Instead, it is important to give adequate consideration to the various aspects of evaluation regardless how they are classified.

The steps of the ex-post evaluation process are:

1. Deciding the evaluation objectives
2. Choosing the evaluation method
3. Data collection:
   - Establishing which data needs to be collected during programme implementation, if applicable in the evaluation method used
   - Collecting data through the monitoring process during programme implementation, if applicable in the evaluation method used
   - Collecting data ex-post if applicable in the evaluation method used
4. Conducting the evaluation and reporting the results
5. Disseminating the results and utilising them in future programme design

1.1 *Evaluation objectives*

Different types of evaluation can take place. In impact evaluation of behavioural programmes the interest is in the effect (change of behaviour) and outcome (e.g., energy saved). Process evaluation refers to “the systematic assessment of an energy efficiency programme for the purpose of improving its design, its delivery, and the usefulness of the quality of services delivered to the consumer” (Spinney et al., 1992).

One should note that full separation of process and impact evaluation can be quite difficult and might not always be feasible. Process evaluation supplements impact evaluation by exploring why savings were achieved. It may include examination of the adequacy of the data needed for subsequent impact evaluations. Another link is the consumer surveys which can simultaneously collect input on programme performance including satisfaction and potential free ridership. Integrating data collection efforts may result in more cost-effective evaluation. (Violette 1995) However, for practical reasons, the two types of evaluation are discussed quite separately in this report.

Below, some examples of evaluation objectives for process and impact evaluation are given.

Examples of process evaluation objectives

- programme context and its significance
- degree of meeting the programme objectives
- analysis of the cost-effectiveness (could also be classified as impact evaluation)
- identification of key actors (stakeholders) and their roles
- strengths and weaknesses in programme design
- strengths and weaknesses in programme implementation (e.g. programme management, coordination and staff)
- identification of barriers to successful penetration of the programme,
- exploitation of programme results and outcomes
- evaluation of non-response (non-participation)

Examples of impact evaluation

- changes in routine behaviour
- changes in investment behaviour
- energy savings or changes in specific consumption and consequent emission reductions
- benchmarking
- market transformation
- calculation of net impact by estimating the impact of free-riders, spill over effect and multiplier effect
- analysis of the cost-effectiveness (could also be classified as process evaluation)

In 25 cases out of 41 it was reported that no process evaluation had been carried out. However, in some of these cases information given included certain elements of process evaluation. It is possible that the respondent was not familiar with the concept of "process evaluation" or did not recognize the evaluation carried out to fall into this category of evaluation. For an example in process evaluation, please see the Norwegian case "Electricity savings in households" (N 3) in Chapter 8.4.

In total, impact evaluation was carried out in 29 cases out of 41. Both qualitative and quantitative impact evaluations could be found in the cases but not surprisingly, qualitative evaluation was more common. Yet, in quite many programmes it also had been possible to take quantitative evaluation to a level where energy savings and/or avoided CO_2 emissions were estimated.

Some of the cases including quantitative impact evaluation aimed at changing only a single variable, e.g. eco-driving (see Nl 9 and UK 5) or use of a particular technology (see A 4, A 6 and N 3). However, examples exist where an attempt has been made to quantify the savings from more horizontal programmes. Here, the challenge is attribution of savings to a particular activity. An UK example is given in Chapter 8.4.

A few examples of evaluation against a pre-formulated market baseline were presented as well. Target group attitudes were plotted in the UK case of eco-driving in Scotland (UK 5) through an on-street questionnaire of commuter drivers prior to the campaign, to achieve a baseline understanding of eco-driving awareness and

behaviour in Edinburgh and Glasgow. These results were combined with results after the campaign. In the German Energy efficiency campaign in households (D 1) the knowledge and attitudes of the public were measured before, during and after the campaign the "before-the-campaign" results forming the baseline. For more details on the German campaign, please see Chapter 8.4.

Evaluation of cost-effectiveness of the programmes was a rarity. This is somewhat surprising given the importance of cost-effectiveness in justifying financing decisions. The likely explanation is the difficulty of quantitative evaluation of the impact of behavioural change programmes.

Typically, not much effort was put in the cases on analyzing the free-riders, spill over effect or the multiplier effect. However, in some cases - particularly those aiming at promoting certain technologies - the multipliers were considered to be important in programme implementation or they might have been one of the primary target groups. This was the case, e.g. in the Austrian programmes promoting heat pumps (A 4) and solar heating (A 6) whereby training courses were organised for installers. In the Dutch EcoDriving campaign (Nl 10) the spill over effect was estimated. In cases promoting certain technologies, free-riders were considered as a potential problem because awareness raising was combined with the provision of subsidies (e.g. "Electricity savings in households" N 3 and in-car devices in the Dutch EcoDriving campaign Nl 10).

Few evaluations were reported regarding non-response, i.e., reasons why certain consumers in the target group participated in the programme or changed their behaviour and why others did not. One reason could be that this was not explicitly asked in the case template. Another reason could be that it is not being systematically analyzed. One example of an analysis of non-response was the Dutch case "Measurement is knowledge" (Nl 2) where it was discovered that 90% of the households could not use an energy metering device provided to them due to poor instruction manual. In two UK cases (National Advertising Campaign UK 1 and Sustainable Energy Network UK 9) low level of participation in one region of the country was recognised to be due to the "unfavourable demographic profile", however, without a clarification why.

1.2 *Process evaluation methods*

Independent programme evaluations typically contain both process and outcome evaluation. The main benefit of independent evaluation is an unbiased view. Self-evaluation - no matter objective or self-critical the evaluator tries to be - cannot reach the level of objectivity and independent evaluation can. The downside of an independent evaluation is, naturally, its higher cost. Examples of both self-evaluation and independent evaluation were presented in the cases. A combination of the two was used in the Finnish "Climate change communications programme" (Fi 3) where the evaluation was first conducted as a self-evaluation by the programme's steering group but and independent evaluation was commenced at the end of the programme.

The approaches available for process evaluation are questionnaires and interviews among the stakeholders (financer, executing agencies and target groups), site visits,

review of programme reports and other deliverables, review of the monitoring results and assessment of the impact evaluation results. These are rather qualitative approaches but more technical process evaluations can be conducted. Technical process evaluations use site visits and surveys to assess the technical aspects of programmes including procedures for selecting programme measures, assessing measure installations, and determining market baselines. Quantitative models can be used for market segmentation and targeting. (Violette 1995)

1.3 *Impact evaluation methods and techniques*

The level of effort put on evaluation in the cases varies considerably. In many cases the reasons for lower level of effort are quite self-evident including available resources, project size and type of activity. In some cases, more effort could have been put on the evaluation activity quite cost-effectively had it been planned from the outset of the programme. For example, in the Finnish Climate Change Communication Programme (Fi 3) it could have been possible to create a baseline though a survey to support later evaluations in a similar way is in the German Energy efficiency campaign in households (D 1).

One categorization of impact evaluation methods is the following: evaluations that use market information and evaluations that use consumer-specific information. Market evaluations can be further categorised into two sub-types: those using national policy models and those using market tracking data that can be gathered through aggregate market analyses. Market-tracking evaluations involve more focused studies of individual markets. Such analyses examine changes in manufacturer, distributor, retailer, and contractor/installer behaviour that could lead to increased adoption of energy efficient measures. The difficulty, however, lies in the separation of programme impact from other development in the market. It may be best suited to provide data for consumer-specific evaluations. Commonly used data in evaluations using consumer-specific information includes billing data, end-use metered data, site data, survey data and programme tracking data (monitoring results). (Violette 1995)

Most of the cases applied evaluations with consumer-specific information. However, examples of evaluations using market information (market-tracking) could be found. Typical examples of the latter were programmes aiming at training salesmen or promoting certain technologies.

Programmes promoting energy efficiency and renewables have most typically been evaluated using the following methods:

- direct measurement (end-use load data)
- billing analysis (energy bills or energy sales data)
- simple engineering estimate (without on-field inspection)
- enhanced engineering estimates (with on-field inspection)

Numerous different techniques have been taken to conduct the impact evaluation (all represented at least one of the case studies):

- analysis of survey results acquired by:

    - mail questionnaires
    - internet questionnaires
    - telephone interviews
    - personal interviews
    - opinion polls
    - consumer panels
    - testing of pupils
    - testing of course participants
    - feedback from course participants
- comparison of “before the programme” and “after the programme” survey results acquired by the above channels
- ex-post survey comparing the target group and non-participant control group
- analysis of survey results with attribution of results to various programmes
- engineering approach combining quantitative monitoring results and default values for savings
- market surveys

1.4 *Evaluation results*

The success of some campaigns was identified with the coverage of the target group by using different communication instruments. The assumption was that a good coverage signifies good implementation of the campaign goals. For example, although a campaign reaches 500 000 children by TV broadcasting that does not tell necessarily about the success of the campaign, if success is measured by the impact. It only reports the media coverage that may be given by media agent. Without feedback systems, that are monitoring and control, the evaluation of real effects of the project remains at a hypothetical level.

The detailed case template included a subjective evaluation of the campaign success (lessons learned). The section, filled in by the programme managers, included questions on how successful the campaign was considered by the financer, implementing agency and the target group. There was some controversy. The official opinion was very often that the project was considered success although no comprehensive feedback was collected. One explanation to the lack of monitoring was often that the project was experimental by nature, a pilot project and was therefore planned narrowly. The campaign designers may think that “doing the right things” (planning) is more important than “doing things right” (implementation). Both of them are of equal importance.

Due to the large stock of information, all evaluation results cannot be summarized here. However, a few examples are given to demonstrate the range of information which can be collected and analysed if monitoring and evaluation are planned from the outset of the programme.

<u>Evaluation of the effect (behavioural change) and outcome (energy savings): Dutch case “Measurement is knowing” (Nl 2)</u>

In the programme, households were provided with a digital plug-in metering device to measure the energy consumption of their appliances. The evaluation

was carried out by internet-based questionnaires. On average, six appliances were measured. 66% of the attending households reported having taken action to reduce their energy consumption. 45% reduced stand-by power consumption and 30% replaced old light bulbs by more efficient ones. Other behavioural changes included reducing the use of a tumbler drier and replacing old white goods by more efficient ones. In outcome evaluation the energy savings were estimated at 250 kWh per household. The outcome evaluation method was not reported but it is assumed to be so-called "simple engineering estimate".

Evaluation of the effect (behavioural change): German case "Energy efficiency campaign in households" (D 1)

The programme is a 6-year (from 2002 to 2008) large-scale information campaign addressing the whole population. The programme results were regularly evaluated by national surveys. The identified changes in public awareness and attitudes between 2002 and 2006 are:

- Knowledge on the cost of stand-by power +4%
- Recognition of the EU energy label +11%
- Use of switched socket extension leads +13%
- Appreciation of the cost savings by compact fluorescent lamps (CFL) +8%
- Knowledge about the variety of CFLs +13%
- Implementation of energy saving measures +5%

Evaluation of the process and impact (energy savings and carbon emission reduction): UK case Energy Efficiency Advice Centres (UK 3)

The objective of the UK Energy Efficiency Advice Centres (EEAC) managed by the Energy Saving Trust (EST) is to help consumers save energy through efficiency measures and thereby reduce their carbon emissions. The two main types of advice provided by the EEACs are verbal advice and home energy check (HEC).

The EEACs have been subject to rigorous monitoring and evaluation. A record of customers is kept in a database. The database is able to identify which customers received which type of advice. The carbon savings impact is assessed through quantitative customer surveys (computer aided telephone interview). The survey is followed by an attribution process whereby the effect of an intervention is assigned to a particular cause or activity. In the case of the EEAC evaluation, it is determined if energy saving actions claimed to be undertaken by interviewees were due to EEAC advice. Information for the attribution is collected by a question in the survey. A reduction is made to account for a proportion of customers who state that they have used more than one EST consumer advice channel. On an annual basis, the savings are estimated at 47 000 tonnes of carbon with the lifetime cost-effectiveness of the programme assessed to be £7.2 tC (based only on the specific funding to

EST). Currently, EST assumes a cautious one year lifetime for behavioural measures; further research in the field has been identified as a key research priority.

Other evaluations carried out include calculation of the cost-effectiveness of the programme and conducting consumer satisfaction surveys.

Evaluation of the process and the outcome (energy savings and market transformation): Norwegian case "Electricity savings in households" (N 3)

The objective of the programme was to reduce electricity use by promotion and subsidising of air-to-air heat pumps, pellet stoves and steering systems for electric panel heaters. An independent ex-post evaluation was performed using mail surveys to households, telephone interviews with technology suppliers and meter reading information. The evaluation consisted of the following elements:

i) Enova's management of the programme
   - documentation of the programme
   - communication with households and interest organisations
   - technology criteria
   - rejections and defaulted grants
ii) Effects among the households
   - description of participating households
   - household satisfaction with the investment
   - energy savings and investment profitability
iii) Effects in the market
   - development in technology markets
   - development in the electricity market
   - social effects and profitability

Programme management got a positive evaluation regarding the application procedure, information and service. Around 50 000 applications were processed with a transaction cost of 10% of the programme resources. About 20 000 households actually received the subsidy of whom 92.1% had installed heat pumps, 6.2% pellet stoves and only 1.7% steering systems.

Free riding was a potential issue in the programme. Overall for the three supported technologies, around 53% of the participants said that they would have purchased similar equipment also without the subsidy. 14% would have purchased a "cheaper" model without the subsidy.

Total annual net energy savings resulting from the programme were estimated at 129 GWh, of which 110 GWh was electricity. The evaluation method used here a combination of "direct measurement" and "simple engineering estimate" as savings were calculated from meter readings and reported changes in other energy use.

1.5 *Reliability of evaluation results*

Little information was given on the reliability of programme evaluation results. Some observations, however, were mentioned regarding impact evaluation:

- UK reported having studied what are the differences between what people report to have done (e.g. in surveys) and what they have really done. Large discrepancies were identified between what people reported and what they actually had done (sometimes by 3 or 4 times) which are then taken into account when reporting carbon savings.

- Some respondents mentioned that in market surveys there can be difficulty to identify autonomous development from the impact of the programme. For example, the Austrian heat pump programme (A 4) faced this problem.

- Attention needs to be paid to sampling in order to avoid errors; the sample (for a survey) should be representative of the target group.

The UK example shows that great care should be exercised when impact evaluations are made using surveys. The respondents may not necessarily deliberately give false information as there is no direct benefit involved; some people will say what they believe is the right thing to do rather than what they actually do. They might also report planned actions. Furthermore, they could report how they perceive themselves behaving instead of how they really do without fully realising the difference. The same phenomenon can be observed in other value-loaded questionnaires. For example, there is usually a big gap between real and reported alcohol consumption.

Sample size is also a key issue particularly where the percentage taking action for each individual measure is a relatively low number and confidence intervals can be large.