

Evaluating Portfolios in Canada: Accepting the Challenge, Negotiating the Minefields, and Achieving Results!

Sohel Zariffa, Hydro Quebec, Quebec, Canada

ABSTRACT

In this paper, we highlight the evaluation planning efforts needed as Hydro Quebec's energy savings objectives rose from 750 GWh for 2006 to 11 TWh by the end of 2015. Faced with such an ambitious target, HQ decided that a rigorous and independent evaluation strategy was necessary.

Having completed our first evaluation in 2006, we have since completed 16 additional program assessments in just two years with nine more evaluations to be completed by year-end. This paper will highlight our experience with Québec research firms, US evaluation consultants, Hydro Quebec's program managers, and our own experience as evaluation managers.

We will discuss how Quebec firms teamed with US evaluation consultants, how language barriers were overcome, how the energy environment within Quebec differs from that of other provinces/states, and how we dealt with these challenges.

The reader will appreciate how a standard procedure for all program evaluations contributed to our success. You will see the template used to standardize the reporting of energy impacts between programs. You will understand that we are now in the process of drafting a set of protocols to be used by all evaluation contractors. We will share with you our evaluation planning process and how it ensures that program implementors track proper metrics and why market characterization studies are important in our efforts.

Our work was complicated by many factors, but we were able to implement a strategy that allowed us to evaluate a complex range of programs in just a few years. This paper summarizes our lessons learned.

The Context

Energy conservation and energy efficiency programs in Quebec and Canada followed a "rollercoaster path" based on the economic, political, and environmental context prevailing at different periods for the last 20 years. In periods of crisis, energy efficiency appeared as the ultimate solution to many societal concerns. Once the crises were over, we returned to our gluttonous energy consumption. Most of what we've built in the first 3 to 5 years was lost in a year or so. It's the price to pay for lack of sustainable commitment to energy efficiency. Let's see if the commitment to energy efficiency can be sustained during this period of recession.

In 2003, things started to change in the province of Quebec. We could sense the wind shift towards a more sustainable approach due to Global Warming alarms, Canada's commitments under the Kyoto Accord, the Clean Development Mechanism, and other triggers favoring longer-term vision regarding energy efficiency. In response, Hydro-Quebec, the public and sole electricity distributor in the province, filed its first three-year DSM plan with the Quebec Energy Board (QEB). The plan targeted 750 GWh savings by the year 2006 with a total budget of \$109 CDN million in support of a portfolio of 16 programs.

These programs did not include any evaluation plans but did include vague program theories. In hindsight, we recognize that these program theories lacked sufficient clarity and proper documentation to gauge program progress or to support the tracking of key program milestones. These programs that were design on a fast-track approach were however cost effective.

In 2008, a new plan called for 4,9 TWh in savings by 2010 with planned expenditures exceeding \$1.18 billion across 27 programs. Recently, the target for 2015 has been set at 11 TWh, the equivalent of the

annual electricity consumption of 407,000 households. Today, HQ disburses an average of half a million dollars in subsidies under its energy efficiency portfolio each working day.

From 2003 to 2005, HQ reported to the QEB some performance indicators such as participant satisfaction, program awareness, and deemed program savings estimates. The estimates were mostly gross savings unadjusted to account for actual net savings documented by a rigorous evaluation effort, as we know it today.

In 2005, HQ recognized that the size of its energy efficiency investments justified a rigorous and independent evaluation. HQ desired an evaluation strategy that would appraise efforts, improve program performance, and establish net program effects.

HQ hired a program evaluation consultant to address impediments common to any utility that chooses to make a serious commitment to the evaluation of their energy efficiency portfolio. A great deal of effort was needed to establish the quality and professionalism of evaluation recently acknowledged by the QEB.

Some procedures and tools were developed systematically while others emerged via trial and error. Where organizational obstacles were present, technical assistance and knowledge sharing was encouraged. Where process-related impediments were found, solutions included the development of guidelines, models and protocols.

How did we start?

Setting priorities within the program portfolio

With 16 programs up-front to evaluate, our first step was to prioritize our evaluation efforts. This exercise consisted of creating a fact sheet for each program. We gathered available information and metrics for each program. This enabled us to see the big picture and assess each program's contribution to the whole energy efficiency portfolio using a value analysis technique. To prioritize each program, we looked at the following items: projected GWh savings, planned expenditures, results from the Total Resource Cost Test (TRCT), the cost of conserved energy, and the status of program implementation. Each variable was weighted by consensus.

Who will do the evaluations?

To rebuild expertise lost after a 10-year relaxation of energy efficiency in the 1990s, we invited Quebec firms to team up with US experts in program evaluation. The list of US experts was drawn from one IEPEC conference participants list. No recommendations or opinions were attached. The list of the Quebec firms was also sent to the US experts.

Examining several forms of Request For Proposals (RFP's) as well as several DSM plans from different utilities across North America, we drafted our own RFP's to solicit bids for evaluation contracts. Our plan was to create a nucleus of expertise in energy efficiency program evaluation in Quebec by selecting three or four consortiums and work with them continuously; feeding them a stream of evaluation contracts.

Sections of the RFP's were translated from French into English and an information session was offered to Quebec firms as an introduction to program evaluation requirements to help them decide whether or not they would need to team with an evaluation specialist. Although we did not require firms to team with a US firm, all chose to do so.

The RFP's approach was characterized by the following:

- Each RFP would cover one or several programs that had similar characteristics, such as addressing the same market, energy efficiency measures, or intervention strategy. This consolidation was necessary to enable us to catch up with the backlog of evaluations.
- Selected firms were retained for a period of three years to enable the evaluator to perform at least two evaluations within the contract cycle. By doing this, the evaluator was in a position to appraise the results of implementing his own recommendations from their initial evaluations.
- The RFP scope of work contained guidelines to present the evaluation activities and all proposals had to follow the same guidelines. This was very helpful for proposal comparison.

The scoring mechanism

Each RFP disclosed the criteria that would be used to assess each proposal. For each RFP launched, a selection committee of five was selected. Each selection committee included two people from evaluation, two from program design, and one expert in market research. Selection team members who were not necessarily the same from one RFP to another, received training on the selection process and the evaluation criteria.

HQ first scored each proposal on its technical merits. A 70 % score for the technical proposal including pertinent expertise and methodology was necessary to proceed to the financial proposal that was submitted in a different envelope. This ensured the cost of the proposed evaluation would not bias the consideration of the technical merits of each proposal. For those who passed the technical evaluation, we opened the financial proposals and the winner was the one who had the highest ‘technical score/financial proposal’ ratio. This enabled us to select the most cost-effective proposals.

We asked firms to relate their experience with program design, development, implementation and tracking, as well as evaluation of energy efficiency programs. Our intent was to ensure that the recruited firm would have field experience. We believe that a good evaluator must also have some practical field experience upon which to draw.

Evaluation Start-up

The next activities were straightforward. We selected the evaluation contractor, provided them with supporting program literature and available program data, and scheduled a kick-off meeting with program managers. At these kick-off meetings, program operators made a comprehensive presentation about the program to the evaluator. The evaluator, in turn, made a comprehensive presentation about the evaluation methodology proposed.

Some important documents and data requested by the evaluators were not always readily available; such as program theory, program logic models, evaluation plans, and fundamental program savings assumptions. For these reasons, some of the proposed methodologies had to be adjusted.

Utilizing Evaluation Findings

Once the final evaluation report was available, program managers were invited to develop an action plan related to the evaluator’s list of recommendations. This action plan was submitted to the QEB along with the evaluation report. HQ explained what they were planning to do with each recommendation along with an implementation timeline. Where recommendations were not appropriate or not applicable for any reason, an explanation was provided to the QEB.

What We Learned and How We Reacted

All stakeholders in the evaluation process learned from this experience. The learning curve was immediately put to good use with the development of several tools. On-course corrections were made to improve the quality of the evaluations, as well as the programs themselves.

1. Our Experience with US Firms

Hydro Quebec Specific Context

Inviting US evaluation firms to team up with Quebec firms was a big success from the start. Some of the top US evaluation firms joined Quebec firms to form evaluation consortiums. US firms found out rapidly that they were dealing with a very particular context in Quebec, a context that needed to be factored into their evaluation methodologies; especially when comparing HQ program results to those in other jurisdictions. The unique characteristics of Quebec that impacted program performance follow:

- Hydro Quebec is a public corporation held entirely by the Quebec government and remains the only major electricity distributor in the province. In 2006, 41 % of the total energy consumption of the province was electricity. HQ does not operate in a competitive market but still has to answer to the QEB. In such an environment, market effects are easier to measure since there is only one major player as compared to a state or province with several electric utilities. Hence, attribution algorithms to establish causal effects are relatively straightforward in most cases.
- A significant proportion of residential space heating (70%) is provided by electricity, mainly baseboard heating. This situation limits the opportunities of new technologies and innovations since electric conversion to heat is close to 100 % efficient at the end use. It also has an important impact in terms of interactive effects. For example, installing CFL's will reduce lighting load but will increase electric heating loads in the winter. Since we have low cooling loads, the savings from lighting measures are seriously diminished when cross-effects are considered.
- More than 96 % of electricity is hydro electric with low GHG emissions. As a result, energy efficiency efforts result in fewer GHG-related benefits when compared to other areas.
- Quebec has among the lowest electricity tariffs in the world: 5.7¢ US per KWh. Low price signal translates into a different appreciation of savings when compared to programs that operate within high tariff jurisdictions.
- Finally, Quebec being a French-speaking province, the language barrier insulates us in term of energy efficiency activities across our borders that are mostly in English. Again, for savings attribution purposes, this isolation is helpful in our efforts.

The organization

Organizational structure also presented challenges for us. Establishing a portfolio evaluation process was difficult because HQ's energy efficiency department is siloed by design component. Each component of program design, development and delivery rests within a different division, making the evaluation planning efforts much more complex since each division has its own perspective and interest. HQ also has limited evaluation staff. In 2006, the Company had only one program evaluator. Today, we have six full-time staff members plus part-time guidance from an evaluation consultant. We are moving from simple survey work to more comprehensive evaluation approaches. As a result, HQ's evaluation department is also on a learning curve.

Double Dipping

Even though energy efficiency activities were on a "Rollercoaster" curve for many years, HQ and

other organizations have continued to promote energy efficiency throughout Quebec. It appears from different surveys that awareness of the benefits resulting from energy efficiency investments remain high. This situation brings us to the question of free-ridership. Some programs have free-ridership estimates that appear overstated based on program activities. We are investigating whether our estimates are failing to distinguish between free rider effects and market effects resulting from previous HQ program activities. Social marketing studies are ongoing to ensure the Quebec context is well understood and factored into the program's evaluation.

2. Our Experience with the Quebec Evaluation Firms

An interesting fact to mention was that initially, no Quebec program evaluation firms were interested in our RFP's; even after being invited several times. These firms were specialized in the evaluation of government programs, such as health or education and were at a loss regarding EE programs. To help close the gap, we included several sections of definition in the first RFP to explain what EE program evaluation was all about. Quebec firms were on a fast track after partnering with US counterparts. The concepts and techniques were completely new to them and most if not all documentation on EE program evaluation was and is still in English.

The most difficult task by far was the evaluation report. Integrating the different sections of a report written in different languages was quite a challenge for the Quebec firms who were responsible for the deliverable. The first versions took up to five revisions just to get it to an acceptable level.

We saw no logic in the report structure. Repetition and contradiction showed a serious lack of project integration. Cut and paste functions were clearly over used and impact results were presented in so many different ways and forms that it was difficult to compare net-to-gross energy impacts from one program to another. Also, causal effects between some of the evaluation findings and recommendations were not always clear.

HQ delivers programs in cooperation with Gas Metro, Quebec's main gas distributor, and the Quebec government's Energy Efficiency Agency. Since HQ took the lead in EE program evaluation, these partners asked HQ to take charge of the evaluation of their programs in partnership. This responsibility contributed to the difficulty in accepting an evaluation report. Every organization had its comments and suggestions that made it even more difficult for a first-time evaluator.

The energy efficiency evaluation field has its own vocabulary that is not always understandable to the average reader. Free-ridership, free-drivers, rebound effects, net-to-gross ratios, etc. are difficult to translate without a glossary that explains each term in the reader's native language. Different evaluations defined technical terms differently. Moreover, for the same definition, often-different terms were used to convey a similar concept. To ensure some consistency, we had to refer to a published glossary.

Translation was another hurdle to overcome. For important meetings such as the start-up meetings, final presentations, or even some key interviews that were overseen by US firms, simultaneous translation was required. Program documents, database content, questionnaires, site-visit protocols, interview guidelines, and even blue prints for new construction programs needed translation. Even though there was nothing we could do to permanently solve that problem, it created some misunderstandings here and there and some serious delays in project delivery. All these elements required some serious coaching by HQ's evaluation consultant.

Several tools were developed by HQ to improve or correct most of the problems discussed above:

- Guidelines to write the evaluation report: This guideline enumerates the main sections that should be included in every report following a standardized report structure: cover page, executive summary, program background, program description, evaluation approach utilized, methodology,

presentation of results for each type of evaluation (process, market, impact), a list of conclusions and recommendations, and ending with a set of appendices.

- A flow chart showing at a glance, the major evaluation activities for a given program. The following illustration shows the researchable issues and data collection strategies; allowing for easy comparison from one proposal to the next and also from one program to another. It also helps remind us of the foreseen evaluation activities for a particular program. This simplification is necessary for HQ's evaluation manager who must manage several evaluation firms and/or programs simultaneously. (Figure 1)

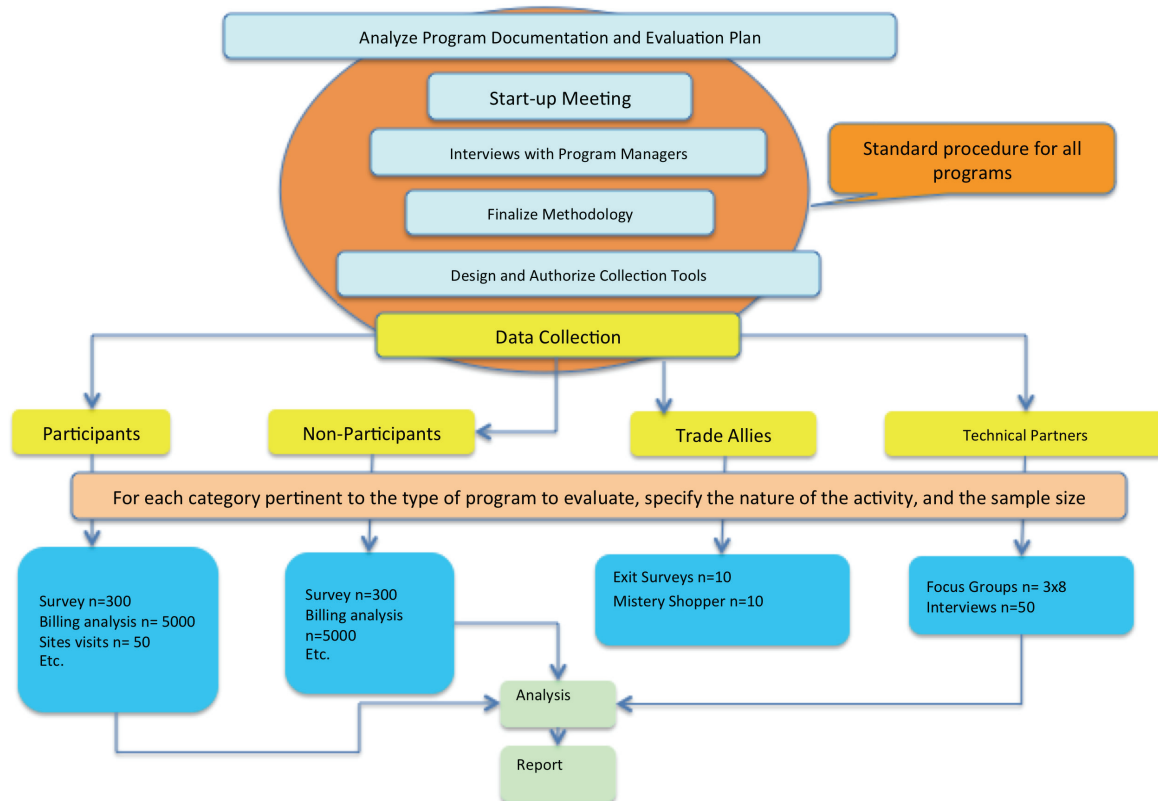


Figure 1: Evaluation Methodology Flow Chart

- A model table to present the impact evaluation results. For resource acquisition programs, the table starts with the number of participants or measures, the gross savings achieved, the different distortion effects (negative or positive), the net savings per year, the life of the measures, and finally the realization rate (results vs. objectives).

The use of the three tools is now compulsory for all evaluators.

Finally, we have to mention that gaining specific expertise in EE program evaluation allowed the Quebec firms to offer their services to other entities that request evaluation services whether they are in the province of Quebec or outside our borders. We have our spillover effects too. We also noticed with great satisfaction that these firms moved from simply noticing some program discrepancies to much elaborated recommendations to improve program performance.

3. Our Experience with HQ's Program Managers

HQ determined that it was important to focus on training its internal staff about evaluation. Staff turnover, inexperienced staff, and new personnel required a better understanding of EE program evaluation.

It was important to highlight in these training courses the fact that impact evaluation is an imprecise science since we try to measure something that is not there anymore. Also, we find that consumer behavior can vary significantly from assumed behaviors towards energy consumption.

The difference between program tracking and program evaluation became clearer to a lot of people over time. Training efforts contributed to this evolving understanding. Early training efforts oriented staff on basic concepts of EE programs. Later sessions related to program evaluation. Offering a session on program evaluation without first providing training about the basic principles of EE program is not recommended. It's like getting your master degree before your bachelor degree. But more important, by understanding how a particular program will be evaluated, program designers paid more attention to the way program offerings were structured.

Fear of evaluation was another challenge to overcome. Program managers always fear being blamed and punished by an evaluator. Even though we tried hard to explain that evaluation is mainly to improve a program performance, human nature cannot be changed overnight. This fear diminished considerably with the approach adopted for the evaluation plan and by our presence when the evaluator interviewed program staff. Program managers who wanted to know the relevance of all the questions asked by the evaluator requested this latter approach.

Because most programs in the early years of the DSM plan were rushed to the market, trial-and-error implementation led to mid-course corrections that added another level of complexity to evaluation efforts especially since tracking systems were not designed with program evaluation in mind. Informative and quantitative evaluations helped greatly to understand the consequences of that type of approach.

Evaluations results were also among the reasons to merge three operational units. Now the design, the development and the marketing units are all part of one operational unit. This vertical integration has resulted in dramatically improved communication between units such that everybody is on the same level of knowledge. These exchanges allow rapid identification of issues that if not corrected, could turn into possible problems when implementing the program.

Market characterizations were not commonly used in program design. Evaluation reports and their recommendations modified this situation considerably. Program managers learned that a market characterization study would:

1. Tell the story of how a market works, in order to help design the most proper marketing strategy.
2. Identify barriers in order to clarify what type of program should be designed.
3. Identify market players in order to select the most appropriate program partners.
4. Document the baseline in order to establish savings objectives net of any natural savings.

These elements were of great importance to the evaluators who tried to justify or appraise the program theory and logic model.

4. Our Experience as Evaluation Contract Managers

Training sessions were offered along with hands-on training. HQ's decision to have an evaluation consultant on-site three days a week with full access to departmental facilities has been beneficial to HQ's evaluation contract managers and has resulted in an easier and more effective hands-on transfer of know-how.

Analyzing several proposals for several programs allowed HQ personnel to build a database of evaluation costs that enable them to better estimate evaluation budgets for future program evaluations.

Because of the particular Quebec context described previously, programs with significant market effects or ultimately market transformation effects needed continuous evaluation efforts and market tracking to establish natural trends in energy efficiency and to better document attribution factors. HQ's evaluation contractors are responsible, if they have the appropriate expertise in market research, to collect the metrics needed for an integrated and continuous vision of an evolving market.

Tools were also developed to help HQ evaluation staff to facilitate their work and to a certain extent, standardize some of their tasks. The most important ones are described below.

An evaluation framework and protocol is being written. Its purpose is to help HQ staff working in EE to better understand program evaluation and its requirements. It is also intended to inform the efforts of the evaluation contractors. The protocol contains 4 modules. It was created by analyzing different well known reference documents¹ and adapted to HQ's organizational structure.

1. The first module provides general information related to concepts and methodologies generally used to evaluate EE programs. The module covers subjects such as :
 - What is evaluation?
 - Why do we evaluate?
 - How do we evaluate?
 - The different types of evaluations
 - The different types of programs
 - The program theory and its logic model
 - The savings assumptions
 - The difference between tracking and evaluation
 - The net to gross concept
 - The evaluation plan
 - Importance of the program database
 - The IPMVP protocol
2. The second module describes the relationship between HQ's evaluation team and the other teams that design, develop, implement and track EE programs. This module provides information that is essential to understand when designing a comprehensive evaluation plan for and with the various team members. For each team, we explain the kind of information needed to build a sound evaluation plan so they can grasp the reason for asking all those questions and what the evaluator is supposed to do with the answers. (Figure 2)

¹1- **Model Energy Efficiency Program Impact Evaluation Guide**. A resource of the national action plan for energy efficiency. November 2007. Steven R Shiller.

2- **Evaluation, Measurement & Verification Framework for Demand Side Management Programs**. OPA version 1.0 and 2.0 October 2007 and January 2009.

3- **Impact Evaluation Framework for Technology Deployment Programs** . An approach for quantifying retrospective energy savings, clean energy advances, and market effects. July 2007, John H. Reed, Innovologie LLC Gretchen Jordan, Sandia National Laboratories Edward Vine, Lawrence Berkeley National Laboratory July 2007

4- **The California Evaluation Framework**. Prepared for the CPUC and the Project Advisory Group. June 2004. TechMarket Works

5- **California Energy Efficiency Evaluation Protocol: Technical, methodological and reporting requirements for evaluation professionals**. April 2006. TechMarket Works

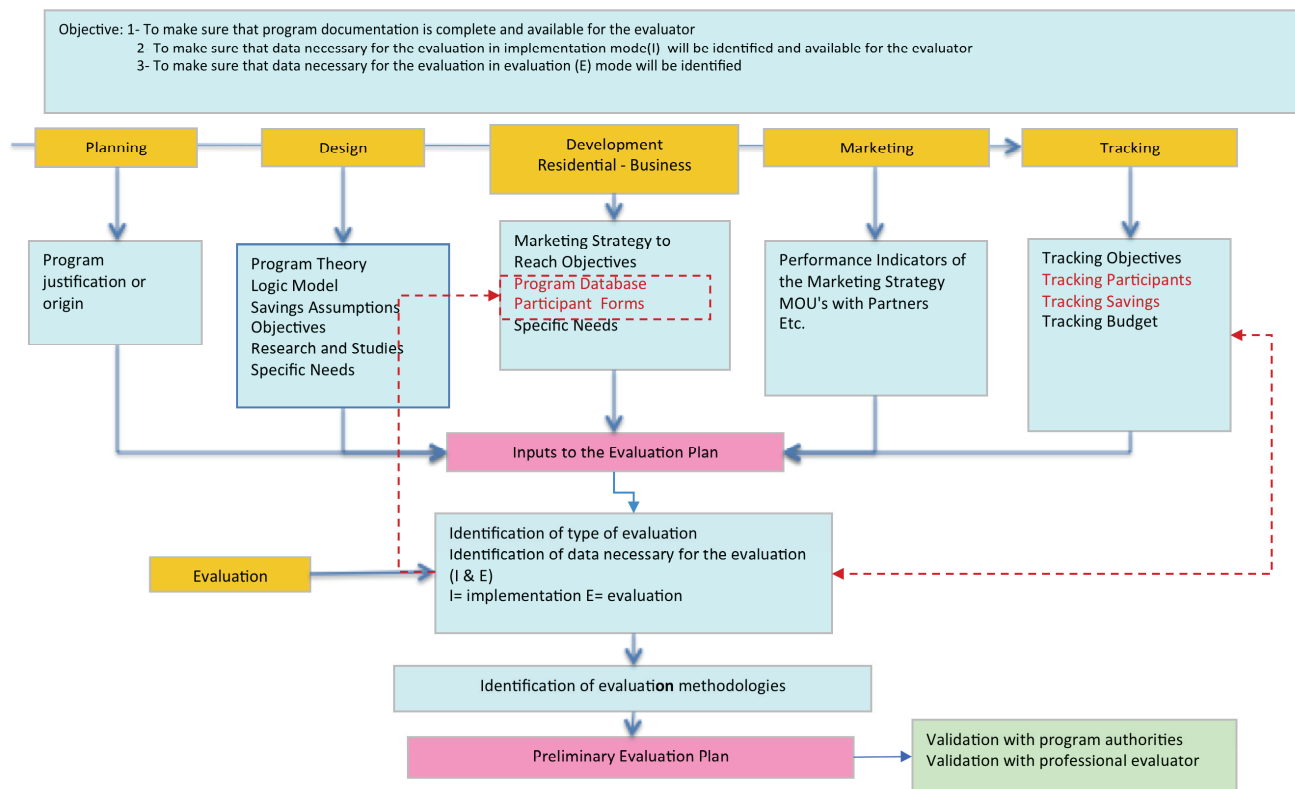


Figure 2: The Relationship between Program Elements and Evaluation Planning

For some utilities, the term evaluation plan refers to the methodology an evaluator will propose to evaluate a given program in response to an RFP. HQ's evaluation plan focuses also on the methodology but with an extra effort to identify data that will be necessary for the evaluator. 'I' type data are those that will be collected during the implementation process via participant application forms. 'E' type data will be collected during MV&E efforts. (Figure 3)

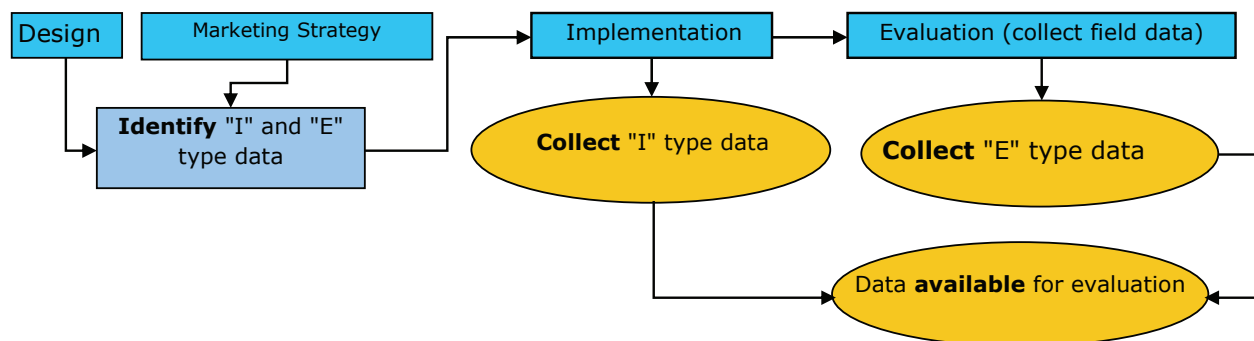


Figure 3: Identify and Collect Data for the Evaluation

3. The third module describes the different tools that the evaluation team has developed to perform its tasks. These tools include report guidelines, forms, impact model tables, evaluation planning templates, etc. These tools are being used at this moment and their effectiveness is monitored to enable us to improve them over time.
4. The fourth module is a glossary of specific terms used in evaluation translated both French and English. Since most of the literature is in English, it was important to translate the English terms into French and to provide a French definition and explanation. This was necessary when we realized that there was a certain confusion in the comprehension of certain terms and that it would be more useful to have a glossary in French then in English to facilitate research.

Analyse de facturation : *Billing Analysis* – L'analyse de données de facturation d'un participant et/ou non participant à un programme pour mesurer l'impact énergétique du programme.

Billing Analysis : *Analyse de facturation* – L'analyse de données de facturation d'un participant et/ou non participant à un programme pour mesurer l'impact énergétique du programme.

After examining a series of evaluation plans from different sources, we established that it would be useful to design our own evaluation plan based on the organizational structure of HQ. Having witnessed what an evaluation looks like without an evaluation plan, we recognized that it was the most important tool to deliver a sound, professional, rigorous and high quality evaluation. Our goal was to make sure that all relative information and data necessary for the evaluation would be identified early in the program development process and made available to the evaluation contractors during their analysis.

The evaluation plan includes a check list section and a narrative section. It was designed so that the first page contains a fact sheet for senior management, while the rest is for specific evaluation purposes. The fact sheet contains information such as:

- Program name
- Program description (short version explaining the what, the how and the who)
- For what type of market or facilities the evaluation study applies (new construction, existing facilities, both)
- Type of program and type of evaluation methods anticipated
- Applicable distortion effects and program net savings objectives
- Timetable and evaluation budget
- Plan version and date² and author

The rest of the plan includes the following sections:

1. Availability of the following support documents for the evaluator with identification of their file name and source for quick reference:
 - Program justification
 - Program theory and program logic model
 - Savings assumptions
 - Marketing strategy
 - Market characterisation study (or any other study)
2. Particular subjects of interest to the program managers to be covered by the evaluator; just in case they were not covered in the list of researchable issues.
3. Details of the business case or the energy savings assumptions

² The evaluation plan that was designed with the collaboration of all other units has three versions. The preliminary version was developed by one of HQ's evaluation team members in collaboration with members from the other teams. The next version was validated by program authorities to avoid any surprises. The final version was validated by a professional evaluator, who in many cases was the same professional who would be evaluating the program.

4. Data needed for the evaluation. This section is in the form of a table and identifies all the data, their sources, the collection method, their type (I or E) and for which type of evaluation they will be needed. The status refers to data that will be collected during implementation “I” (usually through the participant form) and stored in a program database, or collected post-implementation “E” by the evaluator.
5. The foreseen evaluation methods. Based on the preceding information, the method of evaluation will be defined among a series methods commonly used for process, market and impact evaluation. This section also identifies where each method may be applicable (example : participants, non-participants, trade allies, technical partners, etc.) and the sample sizes necessary to reach a specified level of precision.
6. Other evaluation issues: This section in table form, selects a series of issues to be covered in the evaluation such as which distortion effects will be covered, satisfaction and awareness aspects, intentions, reason for participating or not participating, etc.
7. A narrative section on additional information on the evaluation methodologies, if necessary.
8. A narrative section on particular aspects to be monitored during the evaluation because they can affect the outcome of the evaluation.
9. Appendices to the evaluation plan.

As can be seen from the content of such evaluation plans, it has been designed to be used as a mini process evaluation. Missing information or other glitches that could affect program performance would be identified and corrected before launching the program.

Finally, in order to keep abreast of the most recent developments in evaluation techniques and approaches, the HQ evaluation contract administrators started to get involved in the evaluation community and participate in related networking activities. The CEE evaluation committee and the Canadian Demand Side Management Alliance (CDSMA) which gather in Quebec, Manitoba , British Columbia and Ontario are two examples.

Conclusion and Lessons Learned: 17 Programs Evaluated so Far and Counting!

As we mentioned early in this paper, the first HQ DSM plan filed in 2003 had 16 programs in their portfolio. Today, they are running a 27 program portfolio and 13 others are in different stages of design. Our first program evaluation was completed in November 2006. As of today, we have completed 17 evaluations. Five of these evaluations are actually in their second phase while three others have their evaluation plans ready. By the end of this year, we will have evaluated more than 25 programs. Some of these programs are very simple such as prescriptive resource acquisition types of programs while other are very complex market transformation type programs that necessitate considerable evaluation efforts and resources.

The most important lessons learned from the challenge of meeting important energy savings goals were:

1. We need frameworks, guidelines and models for a better management and control of evaluation activities. However, forms and model have their limitations and we have to allow some flexibility and leave some space for free styling.
2. Program evaluation plans should be considered as completed only when the program is implemented.
3. Vertical integration of program activities allows a better control of program activities as well as accountability
4. Market characterization studies are an important tool for both program managers and evaluators.

5. Programs that target important market effects should start tracking market metrics prior to implementation and then on an ongoing basis at regular intervals.

In conclusion, one has to acknowledge that HQ's decision to resort to program evaluation as a means to improve program performance, which is the primary goal of evaluation, was a sound and very profitable decision.