What can we learn from sharing experience about evaluation practices?

Jean-Sébastien Broc, Institute for a European Energy & Climate Policy, France
Gregor Thenius, Austrian Energy Agency, Austria
Dario di Santo, FIRE, Italy
Barbara Schlomann, Barbara Breitschopf, Fraunhofer ISI, Germany
Jamilja van der Meulen, Paul van den Oosterkamp, Energy Centre of the Netherlands, Netherlands
Lovorko Marić, Marko Matosović, Energy Institute Hrvoje Požar, Croatia

ABSTRACT

Many resources are available about state-of-the-art or best examples of evaluations and evaluation guidelines, like in the IEPEC and IEPPEC proceedings. However, what do we know about daily evaluation practices? Is it always easy to find examples of evaluations about a given country, type of policy instrument, etc.? Are evaluations used to improve policies? And finally, how could evaluation practices be improved? This is the kind of issues that the European project EPATEE aims at tackling, focusing on impact evaluations.

About 60 stakeholders from 13 European countries were interviewed or surveyed to better know their priorities about evaluation issues, how they would define the level of evaluation practices in their country and which barriers might impede effective evaluation practices.

In parallel, references were collected and coded to build a knowledge base gathering already more than 170 evaluation reports, papers or guidebooks. Case studies are analyzing more than 20 evaluations to provide concrete and detailed experience feedback about why evaluation is used, how it is performed, what difficulties are encountered, etc.

The objective of the project is not to provide an exhaustive or representative picture of the evaluation practices in Europe, but to gather and develop materials that can be used as a basis for experience sharing activities, as well as to develop an online tool box that will make these resources available in a user-friendly way. The key assumption of the project is that concrete examples and guidance can help overcome barriers that currently limit the use of evaluation. Stakeholders' feedback indeed shows that evaluation can be a very effective tool to improve policies, thereby achieving higher energy savings at lower costs.

This paper presents the results of the first phase of the project, focusing on the main conclusions from the stakeholders' survey, the knowledge base, and the case studies. Feedbacks gathered remind usual no-brainers (e.g., anticipating data collection). It also shows that evaluation is not only a technical issue but that organizational issues (e.g., cooperation between institutions) are critical as well. Learning by doing can help tackle some of the issues (e.g., optimizing data collection), but some issues remain difficult challenges (e.g., getting robust results about net impacts).

Background and objectives of the EPATEE project

International treaties and regulations, as well as national strategies, trigger a great variety of policies throughout countries. The effort put into developing and implementing these policies is well documented for example in existing National Energy Efficiency Action Plans (NEEAPs) in the European Union (EC, 2015; EEW, 2016; NEEAPs, 2017). Emphasis is now increasingly put on finding out how effective these policies have been in view of meeting current and future energy savings targets. European Commission's communications about Smart Regulation indeed highlighted evaluation as a key tool for policy management (EC, 2013). This is also highlighted in the Commission's proposal to recast the Energy Efficiency Directive (EC, 2016): "*In accordance with the Energy Union Strategy and the principles of better regulation, monitoring and verification rules should be given greater prominence*".

Several barriers limit policy evaluation (see for example Broc et al., 2007; Hilden et al., 2014): difficulties related to evaluation issues (e.g., challenges to assess causality or what impacts can be attributed to a given policy), lack of resources, lack of early planning, fear of providing ground for criticism, etc. This results in a lack of quantitative data and impedes evidence-based analysis required to distinguish effective from ineffective policies (Hilden et al., 2014; Huitema et al., 2011). This problem can be tackled by raising the capacity of policymakers and implementers to assist countries to fulfill their obligation and achieve their energy efficiency related targets. The project EPATEE (Evaluation into Practice to Achieve Targets for Energy Efficiency; http://epatee.eu/) provides support to stakeholders both with tools and practical knowledge to make effective impact evaluation an integral part of the policy cycle. EPATEE makes use of existing evaluation experiences in a range of instruments (e.g., EMEEES, 2009; Mundaca and Neij, 2010; Wade and Eyre, 2015). Lessons learned and good practices in how to successfully evaluate the impact and cost-effectiveness of such energy efficiency policies provide the basis for the development of guidelines and good practice evaluation tools.

EPATEE is mainly targeted at raising the capacity of EU Member States to make evidence-based evaluation a part of the policy cycle, and at creating favorable conditions for improving the number and effective use of ex-post impact evaluations of energy efficiency policies.

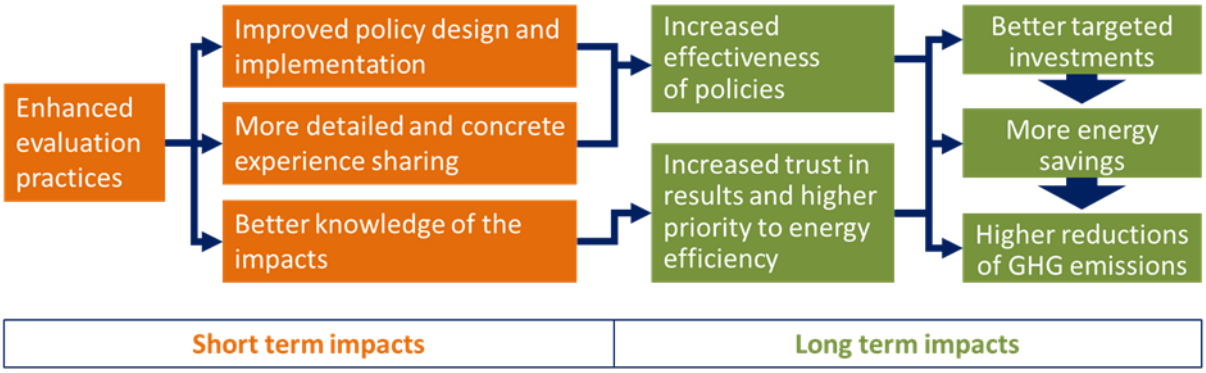


Figure 1. Main project theory of EPATEE.

To achieve this, the project implements several parallel strands, each feeding its results into the other strands and vice versa. This paper presents the methodology used in the project, as well as the first results. A key aspect of the project is to get regular feedback from stakeholders to ensure that project activities and outputs meet their expectations and priorities. These feedbacks are discussed in the next section. The paper then provides an overview of the stock-taking exercise done to collect and analyze resources in order to produce materials for experience sharing. This was done by building a Knowledge Base and doing case studies. Later on, the paper briefly focuses on two key issues raised in the exchanges

with stakeholders: how to evaluate net impacts, and whether evaluation practices should be standardized. Finally, conclusions about these first results are drawn, together with a brief introduction of the next steps of the project (developing an online toolbox and investigating how evaluation can be integrated into the policy cycle).

General feedback from the stakeholders: what do they think about evaluation? what are their priorities?

In order to ensure the effectiveness of the EPATEE project, it was deemed essential to understanding the main issues related to policy evaluation, as perceived by key stakeholders, and to identify their priorities. Besides, stakeholder involvement will guide the design of the tools and activities to be produced over the project duration.

Thus, one of the first actions has been the engagement of the relevant stakeholders – in particular evaluators, evaluation customers, and evaluation users[1] – through two activities:

- interviews with a group of key stakeholders identified at EU level was used to get a qualitative feedback and identify the first needs and priorities for the project activities;
- an online survey among a larger group of stakeholders aimed at collecting a more quantitative feedback, get a better understanding of their needs and rank the priorities identified in the interviews.

Methodology and objectives of the interviews and online survey

The stakeholders selected for the interviews have been chosen considering their role in policy evaluation, either as ministries or agencies involved in such process or as experts in this field. The purpose was twofold: 1) getting a qualitative feedback, as a basis for the further quantitative survey; 2) building contacts for the core of the project community to be further expanded. The selection was thus based on project partners' knowledge of key stakeholders in their country, or based on previous projects or studies. 26 interviews covered 16 EU Member States (Austria, Bulgaria, Croatia, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Lithuania, the Netherlands, Slovenia, Spain, Sweden and the UK). The 30 interviewees were from ministries (12), implementing agencies or authorities (12), experts/researchers/evaluators (3), local authorities (2), and energy companies (1).

The interviews were all done with the same questionnaire including a limited set of open questions. They were carried out mainly as face to face meetings, or through the web or telephone calls when not possible otherwise. A particular attention was paid to avoid guiding answers, in order to get stakeholders' own views. For example, this made it possible to confirm whether stakeholders might have different understandings about what "evaluation" means, and particularly about the frontiers between monitoring, reporting, and evaluation.

The online survey has been promoted to a larger audience of people dealing with energy efficiency policy evaluation, to investigate on a more quantitative basis issues related to evaluation approaches, barriers, and needs. The survey was thus mostly designed with closed questions. Open questions were also used for each main topic, before corresponding closed questions, so that the respondents had first the possibility to express their view without bias and more qualitatively. 36 answers were received, i.e. a 20% answer rate. This was considered a good result as the survey was relatively long

[1] The main target audience of the project are evaluators (i.e. people directly involved in policy evaluation), evaluation customers (i.e. people who commission evaluation activities), and evaluation users (i.e. people who use the results of evaluation, for example for lobbying, research purposes, etc.).

(14 questions) and had to be done in the summer period. Answers came from 13 countries (including 5 countries not covered by the interviews: Belgium, Latvia, Poland, Portugal and the USA). Respondents were mostly from the demand-side of evaluation ("evaluation users": 44%; and "evaluation customers": 24%). 30% of the respondents were "evaluators".

The results and conclusions from the interviews and survey are not meant to be exhaustive nor representative of the evaluation practices of all EU Member States. The objective was to get the views of the key target groups for the project, in order to ensure a good understanding of their needs and expectations. Despite not being representative, the diversity of countries and types of stakeholders provides rich qualitative views about evaluation practices in Europe and made possible to identify key priorities for the project (see analyses below, and the full report: Bini et al., 2017).

How evaluation is perceived: roles and practices

The general concept shared among the stakeholders is that evaluation has to be done in all phases of policy lifecycle: ex-ante, during the effective life of the policy and ex-post. It is fundamental to have data to determine the economic and global impacts of a given policy, thus allowing to understand its results. The ex-post evaluations, on which EPATEE focuses, are important both to evaluate the effectiveness of the policy itself and for planning the new policies (or policy revisions/updates). Some stakeholders also highlighted evaluation as a tool to optimize policy portfolios or to prioritize policy efforts.

An important result of the interviews and survey was the emphasis put by stakeholders on that evaluation allows to understand the impacts of a policy measure, not only in terms of energy savings and costs, but also in terms of additional - positive or negative - side effects (e.g., for competitiveness, fuel poverty alleviation, employment, environmental and social benefits, rebound and free rider effects).

The amount of resources needed to implement evaluation was also reported as an important element. Usually, policy measures with higher resources drain more resources in terms of evaluation, but also the "evaluability" plays an important role. Actually, priority in allocating resources for evaluation is commonly given to measures designed to have direct impacts and/or commit large public budgets (e.g. financial incentives to install energy efficiency solutions), instead of supporting measures (e.g. awareness raising, training activities) that are often perceived more difficult to evaluate and thus less considered. Evaluation priorities might also depend on other factors (such as public debates about the policy, government agenda, or need for reporting obligations). Indeed, most of the respondents mentioned that not all policies were subject to evaluation. Still, systematic evaluation requirements were reported for some countries (e.g., Germany, Sweden, UK). Such requirements were sometimes mentioned to apply mostly for ex-ante evaluation or impact assessment (e.g., Austria), or to be focused on "large budget" policies (e.g., Italy). Some stakeholders also told that the lack of evaluation requirements and/or guidelines was a key reason for the lack of evaluation in their country. Others pointed that evaluating all policy measures does not mean that the measures are all evaluated with the same level of scrutiny. Other factors mentioned about the decision to evaluate a policy include age of the policy, planning cycles, or if debates or critics get higher political or media attention. Moreover, several interviewees raised the difficulty to assess every policy measure individually when measures have strong interactions (e.g., policy packages/portfolios).

Results from previous ex-post evaluations and/or ex-ante evaluations of policies under consideration can inform the design process, leading to optimized measurement, reporting and verification (MRV)[2] and evaluation procedures.

The feedback collected also provided concrete examples where evaluation brings a better understanding of the effects of a policy, particularly about longer-term effects, beyond short-term results registered by MRV. For example, evaluation can show where improvement efforts should be focused, cross effects between policies and energy efficiency measures, take positive interactions in due consideration when designing new schemes or improving existing ones. Some evaluations also brought key evidence to get political support for the continuation, and even sometimes expansion, of programmes (see quotes in Bini et al., 2017). Such examples were also used for peer-to-peer experience sharing through webinars (see recording on https://epatee.eu/events-webinars).

Main difficulties and barriers to the development of evaluation practices

The barriers perceived by the stakeholders are summarised in Figure 2. The three main barriers show a mix of organizational, financial and technical issues:

- insufficient financial resources, for example, due to public budget restrictions and priority given to funding implementation;
- lack of interest from policymakers and public managers, i.e. a cultural barrier that exposes Member States and local governments to an ineffective use of the available resources and reduces the possibility to learn by doing, and that might be for example due to a fear that evaluation finds the policy less effective than planned;
- lack of reliable data to evaluate non-energy effects (i.e. important aspects and impacts of policies are not covered by the evaluation process, as already mentioned in the previous section).

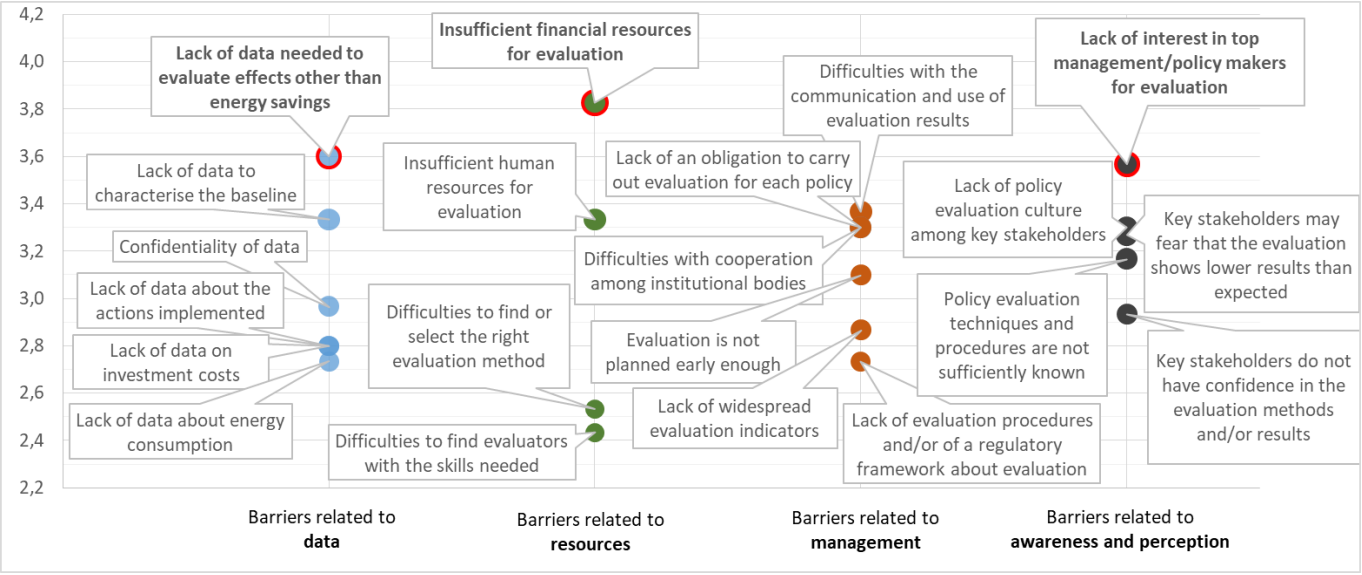
The two first barriers are also highlighted in the literature (see e.g., Broc et al., 2007; Hilden et al., 2014). The third one (and its specific link to non-energy effects) seems to have received more attention in the recent years (see for example developments in other European projects such as ODYSSEE-MURE or COMBI). While data problems have always been pointed by evaluators as a critical issue for any evaluation. It is also interesting to note that the grading of the barriers (5-point scale) was similar between the three profiles of respondents (evaluation customers, evaluators and evaluation users). Overall, evaluation customers used slightly lower grades, but resulting in the same ranking. The difference was, however, higher for the grades about barriers related to data issues. Evaluation customers would thus perceive these issues as a bit less troublesome, particularly compared to the perception of evaluators. This result is based on small sub-samples, so this assumption would need to be further tested.

Another key barrier to evaluation pointed by some interviewees is the lack of trust that stakeholders may have in evaluation results. This is indeed essential for policymakers and other stakeholders to take these results into account. Trust may depend on how stakeholders perceive the quality of evaluation whether they were involved in the evaluation process and whether results are transparent.

To conclude, the lack of interest in evaluation sometimes shown by the top management and the fear to see results less good than expected can both explain the lack of priority/resources dedicated to

[2] Stakeholders emphasized the importance of distinguishing between MRV and evaluation processes. MRV provides data and feedback on a regular basis for managing policy schemes. Evaluation provides an in-depth and possibly independent analysis of the schemes and their impacts, in order to verify the cost-effectiveness of policies, identify the effects on the market, draw recommendations for enhancing current policies and establish new ones, etc.

evaluation. This may also explain why some respondents consider that the lack of an obligation to perform evaluation can be a major barrier. On the other hand, if there were a stronger support from the top management to do evaluations, there would be no need to push for an obligation to evaluate.



Note: scale from 1 (barrier with low influence) to 5 (very important barrier)

Figure 2. Question "*Please grade the importance of the barriers to evaluation*". *Source*: Bini et al., 2017.

Lessons learned about overcoming the barriers and developing evaluation practices

The interviews and the open questions of the online survey made it possible to identify the following lessons learnt from stakeholders' feedback, about why evaluation was important and how evaluation practices could be improved:

- without evaluation, it is not possible to know the effects of a policy, and therefore to make decisions based on evidence and to support policymakers' decisions (in other words, the question might not be "how much does evaluation cost?" but "how much does it cost not to do evaluation?");
- evaluation should be included from the planning phase of the policy instrument, identifying data needed, people in charge, MRV, procedures, etc.;
- integrating evaluation from the start of the policy process makes it possible to get more reliable results, and to optimize data collection;
- policies should not be evaluated as a single measure, but considered in the overall framework of energy efficiency policies, which often implies interactions between policy measures;
- information flows between policymakers, the policy measure and evaluators should be well planned for the evaluator to understand the instrument and the underlying reasons that led to certain policy design decisions.

Sometimes evaluation allows highlighting underestimated, or not foreseen, effects related to the implementation of policies. For example, in Croatia, the installation of individual heat cost allocators in multifamily buildings showed costs higher than the savings depending on the type of buildings. In Amsterdam, the actual energy savings related to grants given on the basis of the building's energy labels

improvements were lower than expected based on the modelled differences in energy consumption between energy labels. Collecting and sharing such examples is a way to raise awareness about the added value of evaluation, as discussed in the first EPATEE webinars (see https://epatee.eu/events-webinars).

Defining and adopting evaluation guidelines and/or requirements was also mentioned in the interviews among possible solutions to tackle barriers to evaluation. 9 interviewees (out of 26 interviews) thus highlighted that standardised methodologies, procedures, indicators and/or guidelines would be useful or even needed, to improve the evaluation practices and/or to ensure a larger comparability among different policies and countries. But 4 interviewees also raised a concern that evaluation needs to be tailored to the policy analysed, taking into account its objectives, as well as the national background (including national evaluation culture or practices). This debate represents one of the sensitive challenges to deal with to improve evaluation practices. Combining both views could be done by providing standard concrete guidelines for evaluation, but leaving each stakeholder free to choose if and how to use such guidelines. This is for example the approach taken by the World Resource Institute's Policy and Action Standard for GHG measures[3]. One interviewee also suggested that more experience sharing between countries would help to harmonize evaluation practices on a voluntary basis.

Building a Knowledge Base for evaluation of energy efficiency policies

Methodology and objectives of the Knowledge Base

The Knowledge Base developed within the project encompasses evaluation reports on energy efficiency measures, guidelines on evaluations of energy efficiency measures, meta-analyses and papers describing, discussing and analyzing respective evaluations. The goal is to provide the concrete materials and information to be used for the development of an online toolbox (see Figure 5, at the end of the paper). The toolbox and thus the Knowledge Base aim to reduce the inaccessibility of information of energy efficiency evaluations. We perceive inaccessibility due to missing information on existing good evaluation reports and practices but also due to lack of clarity on information. There is a huge variety and heterogeneity of evaluations of energy efficiency policies. This heterogeneity is grounded on the different policy measures, sectors, actors, countries of evaluations and methods that are applied in the different studies. Moreover, the available literature for energy efficiency evaluation is diverse and may follow different evaluation methodologies depending on the available resources and knowledge. Therefore, the Knowledge Base strives to facilitate the access to information on evaluations by offering a kind of structure or classification of evaluations, which is reflected in the search structure of the toolbox. Ultimately, the Knowledge Base strives to contribute to a better policy design of energy efficiency measures by providing information materials, e.g. on robust evaluation methodologies.

The Knowledge Base has been developed iteratively based on initial literature scoping and discussions within the project team. A clear template for the scope of the appropriate literature has been laid out to ensure both, the robustness of included sources and diversity of relevant information. The goal of the Knowledge Base is to give a cohesive overview of evaluations covering as many EU member states as possible (i.e. also including references in national languages), addressing the main types of policies, methodologies, and sectors, and last but not least include relevant evaluations analyzing major policies in the respective country. The classification of methods and policies in the Knowledge Base is based on the classifications applied in the Odyssee-Mure project (http://www.odyssee-mure.eu/; Schlomann and Eichhammer, 2011). A primary review process was undertaken, which focused on highly cited papers from Google Scholar, Web of Science, and IEPEC&IEPPEC papers (http://www.ieppec.org/proceedings/). A

[3] https://www.wri.org/sites/default/files/Policy_and_Action_Standard.pdf

second review process was based on suggestions from within the project team, which come from nine different EU countries. The selected evaluation studies are primarily ex-post evaluations with a particular focus on energy savings. However, further benefits such as reduced GHG emissions or economic impacts are considered as well.

Overview of the current content of the Knowledge Base

Currently, the Knowledge Base comprises about 170 entries. They comprise information about the type of evaluation (i.e. evaluation report, paper, methodological paper, meta-evaluation and guidelines), type of policy instrument, sectors, geographical scope, language, year of publication, type and objective of evaluation, data collection and calculation methods as well as on the baseline scenarios, normalization and adjustment effects and impacts apart from energy savings.

The type of studies comprise empirical (evaluation reports and papers) and analytical (guidelines, methodological, meta-evaluation papers) documents. The bulk of evaluations come from the UK. This is grounded on the fact that in the UK, evaluations of policies have been part of the policy package for many years and that searches were first made with keywords in English. But overall 28 countries are represented in the base (including 12 non-EU countries). Similarly, there are entries about all sectors (residential, services, industry, transport, agriculture, but there is a strong bias towards the residential sector, which typically employs financial or fiscal support measures. While agriculture gets little notice. This might be explained by the priorities in national energy efficiency policies, as shown in the NEEAPs.

This distribution focused on specific countries or sectors or policies, restricts the possible conclusions that may be drawn regarding coverage or usage or practice of evaluations. The objective was to collect and make available material for experience sharing and capacity building, not to be representative nor exhaustive.

Given the selection focus, not surprisingly, most of the studies are ex-post impact evaluations, often with bottom-up calculation methods. However, there is a good level of precedent information regarding the various methodological issues, which are applied with homogeneity in both analytical and empirical papers (see Figure 3). Temporal analysis suggests that correction factors that have a precedent body of information within the analytical papers (e.g. on spillover, rebound or free-rider effects) have their methodologies applied within the empirical evaluations (see Figure 4). When there is a low level of precedent methodological information (additionality, energy security) these additional evaluation criteria are more commonly omitted from evaluation or apply less homogenised methodologies.

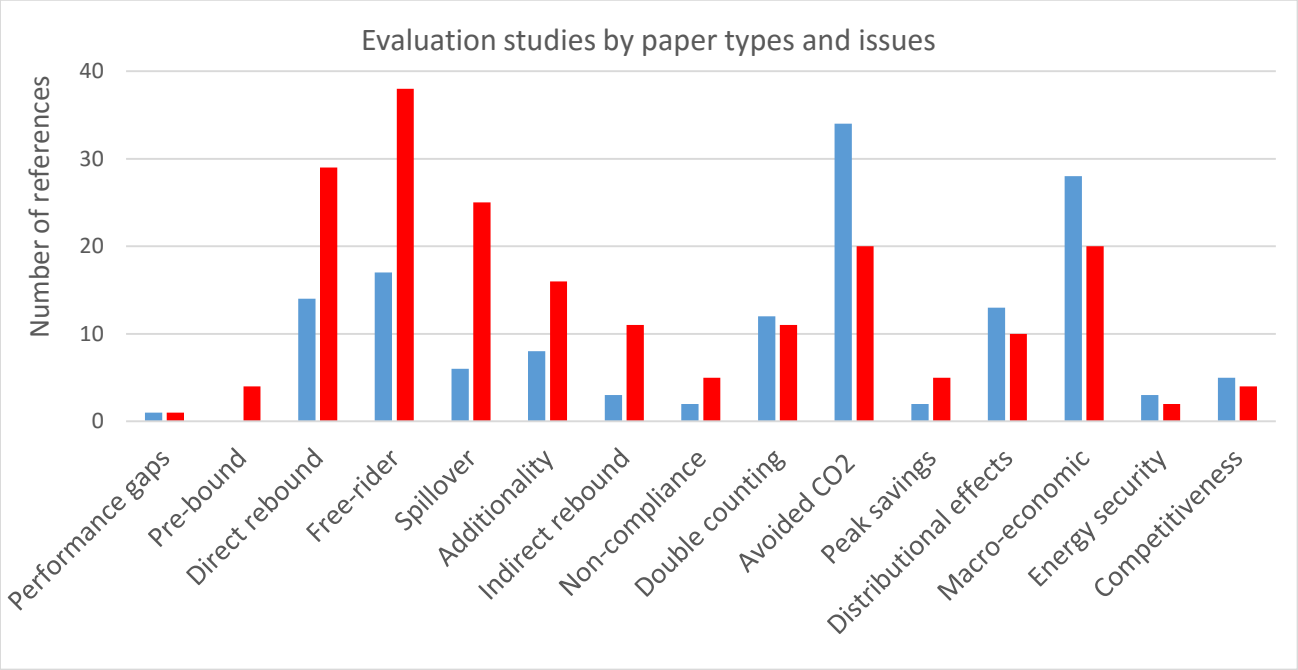


Figure 3: Distribution of empirical (in blue) and analytical (in red) papers by evaluation issues

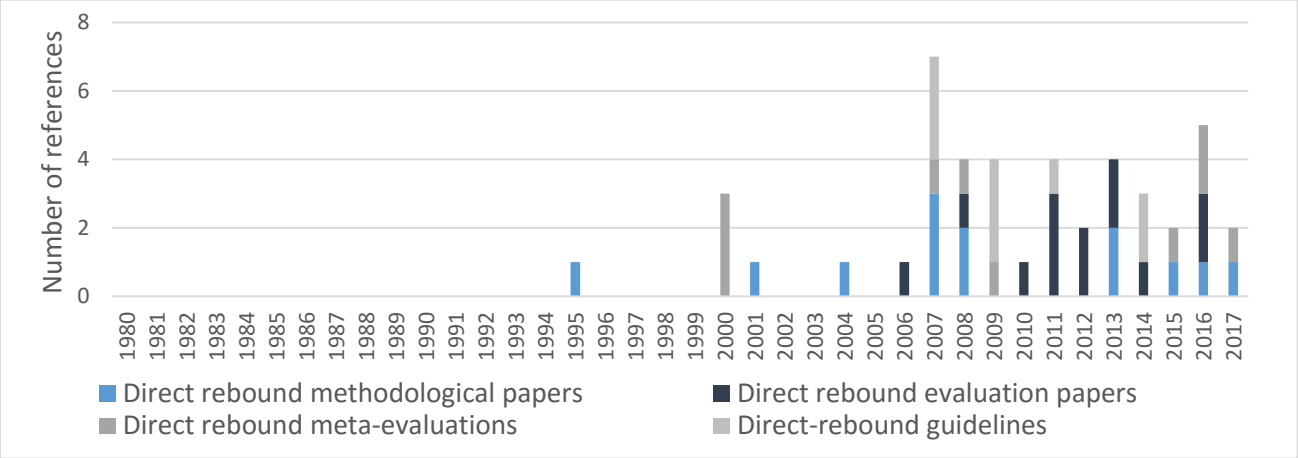


Figure 4: Distribution of papers discussing rebound effects

Overall, energy savings are covered across all evaluations in more or less depth, and methodologies are applied appropriately across all studies. But they are somehow measure-type or sector dependent, i.e. evaluations of fiscal policies apply rather top-down methods while evaluation of financial support employs rather bottom-up methods. However, the framing of the Knowledge Base in particular with respect to methodologies means that levels of information might be lost regarding the specific methods applied. Particularly because when a study includes the use of a combination of methods or a specific type of methods (not included in the types predefined in the Knowledge Base and MURE database), it is then sorted in the category “other methods” (see Table 6 in Annex).

Making the Knowledge Base a living resource

All the information of the Knowledge Base will feed-in an online toolbox (see further down), which will be accessible through the EPATEE internet page (https://epatee.eu/main-results). Users can do a direct online-search in the Knowledge Base on the same internet page as well. The online search of the Knowledge Base will allow searching in both a simple and advanced type of search. A simple search can be made by type of study, policy instrument, sector, and geographical scope. An advanced search is feasible allowing for searching by all criteria. Multiple selections within a category are possible. That is, a user can search for more than one policy type or sector. In addition, the studies are linked to the case studies (see next section). The output of the online search contains all information on the criteria that are selected as well as the title of the study, the internet address if available, and the study as pdf if available and publicly accessible.

To keep the Knowledge Base, and hence also the toolbox, updated, it is considered a living document. This means that further or new studies, which are recommended by users will be added on an annual mode by the project team. Before the update, the project team briefly reviews the recommended evaluation studies.

First lessons learned from case studies

Methodology and objectives of the case studies

The first motivation to do case studies about evaluations was the observation that experience sharing about evaluation practices is often limited due to the lack of time for stakeholders to disseminate or document evaluation works, and due to the many languages in the European countries. Front-runners and researchers may publish their work in scientific journals or international conferences (see for example the review done by Wade and Eyre, 2015), but most of the "regular" evaluation works remain in national language and are not always easy to find. Therefore, actual evaluation practices of the stakeholders and the difficulties they encounter when doing evaluations are not well known. Moreover, evaluation results are sometimes disseminated without detailed explanations about their evaluation methods. This creates limitations for experience sharing, as well as for a correct understanding and use of evaluation results, as also noted by Haug et al. (2010) about climate policies. This is particularly true for ex-post evaluations.

The objectives of the case studies are therefore to analyze concrete examples of evaluations, with an emphasis on why evaluation is used, and how it is performed. The aim is that these case studies form materials for peer-to-peer experience sharing, and help making information more accessible, providing data as transparent as possible.

Like for the interviews and survey, the selection of the case studies was not meant to be representative of evaluation practices in Europe, but to cover a diversity of situations (policy instruments, sectors, countries, evaluation methods) and to provide interesting experience feedback in order to identify both, good practices but also difficulties encountered.

The analyses combine two sources. First, the evaluation report and related information available (online or in paper documentation) are reviewed to fill in a template (see table below). Second, an interview is made with the evaluation customer and/or the lead evaluator to validate this analysis and get a complementary and direct feedback about the evaluation (reasons for the evaluation, lessons learned about the policy, lessons learned about evaluation practices, etc.).

Table 1. Sections included in the template for the EPATEE case studies.

1) short description of the policy measure	3) main data on energy savings	5) insights about other aspects monitored/evaluated	7) interview with evaluation customer or evaluator
2) main data on means and outputs	4) short description of the evaluation method	6) focus on a particular evaluation issue/practice	8) references

When presenting data about energy savings, costs, etc., particular attention is paid to explain the corresponding unit and scope. One frequent observation about data available in reports or online is that part of the information is implicit. This might create confusions: for example, it is not always specified if data are about final or primary energy, what period is taken into account, what types of costs are included.

In addition to case studies on concrete examples of evaluations, synthesis papers are also being prepared on evaluation issues that were noted as priorities by stakeholders (see interviews and survey above), and that represent key challenges in terms of evaluation practices. These synthesis papers combine a concise literature review and concrete examples collected in the case studies.

Table 2. Sections included in the template for the EPATEE synthesis on key evaluation issues.

1) Scope and definition of the issue	3) Common approaches used to tackle the issue	5) Bibliography
2) Insight from the literature	4) Concrete examples	

Overview of the case studies

Table 3. List of EPATEE case studies.

Country	**Sector**	**Type of instrument***	**Name of the policy measure**
Austria	Industry	Financial	Environmental Support scheme (Umweltförderung im Inland)
Austria	Transversal	Policy mix	City Energy Efficiency Programmes of Vienna
Belgium (Wallonia)	Residential	Financial	Primes Energie (grants for energy renovation)
Croatia	Services	Financial	Energy renovation of public sector buildings
Croatia	Residential	Information/Education /Training	Individual heat cost allocators in multifamily buildings
Denmark	Transversal	EEO**	EEO scheme
Finland	Industry and services	Co-operative measure	Energy Efficiency Agreement for Industries
Finland	Services	Information/Education /Training	Voluntary audits for municipalities
France	Residential	Financial	Tax credit for renovating dwellings
France	Transport	Co-operative measure	Voluntary agreement for freight companies
France	Transversal	Financial incentives	"Future Investments" programme
Germany	Industry and services	Information/Education /Training	Learning Energy Efficiency Networks Initiative
Germany	Residential	Financial	KfW programmes for buildings
Germany	Transversal	Financial	Energy Efficiency Fund
Greece	Services	Financial	Subsidy scheme for local authorities
Ireland	Residential	Financial	Better Energy Homes
Italy	Residential	Financial	Tax credit for renovating dwellings

Italy	Transversal	EEO	White Certificates Scheme
Lithuania	Residential	Financial	Renovation programmes with EU funding
Netherlands	industry	Co-operative measure	Multi-year agreements in the industry
Netherlands	Residential	Financial	Subsidy scheme for housing corporations in Amsterdam
Nordic Countries	Residential	Legislative/normative	Nordsyn (market surveillance)
UK	Residential	EEO	Supplier Obligations
UK	Residential	Financial	Warm Front
US	Energy sector	Capacity market	Auctions for capacity markets
US	Residential	Financial	Weatherization Assistance Project

* typology taken from the MURE database, see (Schlomann and Eichhammer, 2011).
** EEO: Energy Efficiency Obligation scheme
Note: complementary information about the type of methods and evaluation issues included in the case studies can be found in Table 4 in the Annex of this paper.

The first EPATEE synthesis papers are about:

- Assessing net energy savings;
- Linking M&V tools and evaluation practices;
- Investigating the differences between energy savings based on engineering estimates and energy savings based on metered data.

Case studies and synthesis papers are available at: https://epatee.eu/case-studies

Key messages from the case studies

The case studies show the diversity of evaluation practices. The scope, focus, and depth of evaluations depend on evaluation objectives: what the evaluation customers want to know/investigate, what they need in terms of reporting, what the evaluation will be used for, etc. Evaluation customers and evaluators pointed out the importance of defining clear evaluation questions and scope. For example, the evaluations of the Danish EEO scheme had a clear focus on assessing the additionality of its impacts. Whereas the evaluation of the subsidy scheme for housing corporations in Amsterdam raised the question of the uncertainty about energy savings estimated based on Energy Performance Certificates. This confirms that care should be taken when comparing results of different policies and/or evaluations (see also Broc et al., 2017).

Practical issues might create barriers or difficulties to evaluations. Several examples illustrate that when data collection is not prepared early enough, this generates extra costs or even missing data. A good practice to overcome this problem is to create a legal framework for data collection, for example, to include conditions in financial schemes for participants to agree with providing data for the evaluation. This might however not be sufficient to solve all problems, as for example the frequent difficulties to collect reliable data about the situation before renovation works. Another usual practical problem is incompatibilities between databases that need to be combined. This is often solved by using feedback from the evaluation to improve the monitoring systems. Most of the case studies indeed emphasized the key role of monitoring systems in making evaluation feasible.

Finding the right timing for the evaluation is not an easy task. Evaluators often report that they have to produce results in a short time-span. Decision-making timeline is rarely the same as the time

needed for impacts to be visible. This often adds to limited resources to collect data. Evaluators then need to choose pragmatic approaches, which might be subject to criticism by stakeholders not satisfied with the results. A good practice to tackle this issue is to plan evaluation on a regular basis. In this way, problems encountered in the previous evaluation can be taken into account to improve the evaluation approach/method and get a consensus among stakeholders about the reliability of the results.

Evaluating policy impacts requires a good understanding of the background. Evaluation customers mentioned the importance to ensure that evaluators can have access to or collect not only data needed to evaluate the impacts but also the knowledge of the policy background. This avoids analyses or recommendations that are meaningless because out of context. A good practice is to keep track of the policy history. As policy officers might change over time, this avoids losing the policy memory. When not done on an on-going basis, gathering all the relevant information about the policy and its changes can be very time-consuming when preparing the evaluation.

Policy officers should not fear evaluation but use evaluation to build trust among policymakers and stakeholders. The fear of bad results is often reported as a barrier to evaluation. At the opposite, feedbacks show that a transparent monitoring & evaluation process is an effective way to build trust and increase the legitimacy of a policy. Gathering evidence will improve the quality of the debates about the policies. Evaluation customers often mentioned that the monitoring & evaluation helped policy officers to get political support for their policies (see for example Energy Efficiency Agreements in Finland, or Better Energy Homes in Ireland).

Focus on two issues raised in the first European EPATEE workshop

Issue 1: how to evaluate net impacts

Analyzing the net effect of policies is one of the main challenges in impact evaluation. Ideally, an impact evaluation shows only the effects that were triggered by the respective policy, separating them from other effects (e.g. other policies, external factors, trends). In practice, this ideal situation can only be approximated, particularly because it is rarely possible to have a relevant control group (and to have data about it). This is confirmed by the case studies. Discussions at the workshop covered the concepts of additionality, free-riders, and spill-over effects.

The concept of **additionality** serves at identifying net effects but is defined differently across Member States and policies. The main challenge is to set the baseline, which often cannot be clearly defined because it may differ from one evaluation to another, depending on many factors including the evaluator, the time period and other factors.

Assessing the extent of **free-riders** in policies helps to come closer to net effects. Certain programs can be assumed to have either no or strong free-rider effects. For example subsidies on boilers less than 15 years old should have limited free-ridership because the lifetime of the appliance is 25 years. On the other side, high free-riders can be assumed for double-glazing as this action is mainly implemented for non-energy reasons (e.g., aging, noise or aesthetic issues). In this case, higher energy efficiency can be a side-product. Ideally, the assumptions made need to be tested or evaluated ex-post. But feedbacks show that this remains often challenging to do in practice: market data are rarely available to make rigorous statistical analyses possible, and surveys include several risk of bias (e.g., social desirability).

These observations lead to the importance to complement the analysis of free-rider effects with **spill-over effects** – indirect positive/enhancing effects of a policy. The case studies show that these effects are rarely evaluated, mostly because this would require additional data collection that cannot fit in the evaluation budget and time-span.

In any case being very precise about all these adjustments often comes with bureaucratic hurdles preventing people from using the program. Denmark pursues a pragmatic approach towards net effects.

Instead of trying to identify all single factors influencing net effects, the targets for the EEO scheme are increased regularly, and actions for which evidence shows low additionality are credited with fewer energy savings, or even removed from the scheme. The additionality has been evaluated with surveys that provide a basis for the discussions between the public authority and the stakeholders. This approach is also easier to communicate. However the latest evaluations have shown that more verifications and controls were needed, for this approach to remain effective when targets are more difficult to reach. The Danish Ministry therefore increased the resources for these verifications.

Finally, it has to be noted that policies are usually part of a broader energy or climate strategy thus embedded in a policy mix with overarching energy efficiency targets. Calculating net energy savings for each policy instrument would lead to the danger of not taking into account synergies between policies, finally leading to an under-estimation of the effects.

Issue 2: whether or not to promote standardization of evaluation practices

The issue of standardization or harmonization of evaluation practices was not initially included in the questions of the stakeholders' interviews (see above) nor in the agenda of the first EPATEE workshop. However, it came out in the discussions about barriers and evaluation costs and raised debates.

Arguments in favor of harmonization were about ensuring the quality of evaluations, or the needs for comparison purposes. The EU Energy Efficiency Directive indeed resulted in a higher interest in comparing results between countries and/or policies, particularly due to its Article 7. Previous projects and studies (see e.g., Broc et al., 2017), as well as discussions at the workshop, proved that it is currently impossible to directly compare results reported by the Member States to the European Commission, without entering in careful analysis of these results and how they were calculated.

One question is whether it would be possible that all Member States report energy savings in a harmonized way that would make comparisons easier. It was suggested that one possibility could be that each Member State would be free to use their own way to account for energy savings when dealing with the policy at the national level (for example because special rules may be applied due to policy priorities), but that the results would then also be calculated according to European guidelines when reported to the European Commission. However, it was explained that first, this would create an additional administrative burden for the Member States. And second this would not be technically feasible, as this option can only be possible if all Member States collect at least a same minimum dataset. In practice, the collected data differs from country to country and even from one policy to the other within the same country. This makes it impossible to use the same calculation methods everywhere. Several attempts to promote harmonization were done in the past (see e.g., EMEEES, 2009 and the resulting recommendations issued by the European Commission). But none of them succeeded in becoming the common practice.

One alternative could be to promote harmonization of the way to document energy savings. This is, for example, the approach used in standards developed about energy savings calculation (first at the European level, CEN standard EN 16212, and now at the world level with ISO standards, like the upcoming ISO 50046). These standards may not always be applicable for evaluations at the national level, as most of them are meant to be used by companies, for example in line with ISO 50001 about energy management. But the use of clear documentation guidelines could be an interesting alternative to explore. Harmonized documentation of energy savings would not only be useful to facilitate comparisons and benchmarking. It would also be useful to document the evaluation approach. When results are not well documented in reports or other information sources, the memory is lost over time as policy officers change.

About the issue of ensuring quality and improving evaluation practices, there were debates about whether harmonization of evaluation methods can be the most appropriate approach. Some participants were more favorable to promote experience sharing, as done in the EPATEE project.

Perspectives and upcoming work

Main conclusions from identifying good practices and gaps

One of the following major activities of the EPATEE project is to summarize the lessons learnt from the knowledge base and case studies. This synthesis report will also take into consideration the results from stakeholders' feedbacks (see above). It will include analyses on the geographical and thematic scope of the EE policy issues, applicability from lessons learnt from one country to another, definitions of indicators (technical, economic, environmental), calculation methods used and adjustments taken into account, differences between the recommendations from the literature and current practices analysed in the case studies.

The goal of the research would be for a future evaluator to use this summary as a basis for own research in a respective country before the evaluation takes place, as a kind of check list. Based on the frequency of researched issues, the evaluator can choose a set of criteria within a policy cycle for comparison points and suggestions for further stages of the evaluation cycle. The criteria would be divided into technical (e.g. correct usage of calculation methods, adjustments for performance gaps), economic (cost-effectiveness of policies and programmes, investment triggered...), environmental (CO_2 emissions, health, and air quality benefits) and societal (satisfaction with the scheme, number of educated workforce per saved kWh...), and whether these were used (correctly) as indicators for the success of a respective energy efficiency policy, and how those missing could be used in future evaluations, if applicable.

Building on the materials gathered to develop an online toolbox

One of the upcoming activities in the EPATEE project is to provide an online tool that should help the users (policymakers, evaluation customers, evaluators and other stakeholders) to find an appropriate evaluation approach that meets the requirements of these users. The focus is primarily on ex-post evaluation/realized savings, and not on ex-ante/expected savings.

Important elements that will determine the final characteristics of this online tool include:

- The needs of the stakeholder(s), from interviews surveys and workshops
- The available evaluation methods, with their resources needed to implement the method and the results that the method can deliver
- Background information on evaluation (the Knowledge Base)

The possible issues for users of the online tool are sketched on the left side of Figure 5 below. These issues need to be translated into an advice. In most cases, this will be a specific methodology (the green box of the picture). This advice makes use of documents and references that are part of the Knowledge Base (KB), case studies and synthesis produced within the EPATEE project (see previous parts of the paper) or existing references on standard methods and evaluation protocols or guidelines. In the online tool, we will focus primarily on impact of energy savings. Secondary aspects like Non-energy Benefits (NEB), for example, employment effects, energy poverty, and cost-effectiveness will be treated as other needs of stakeholders.

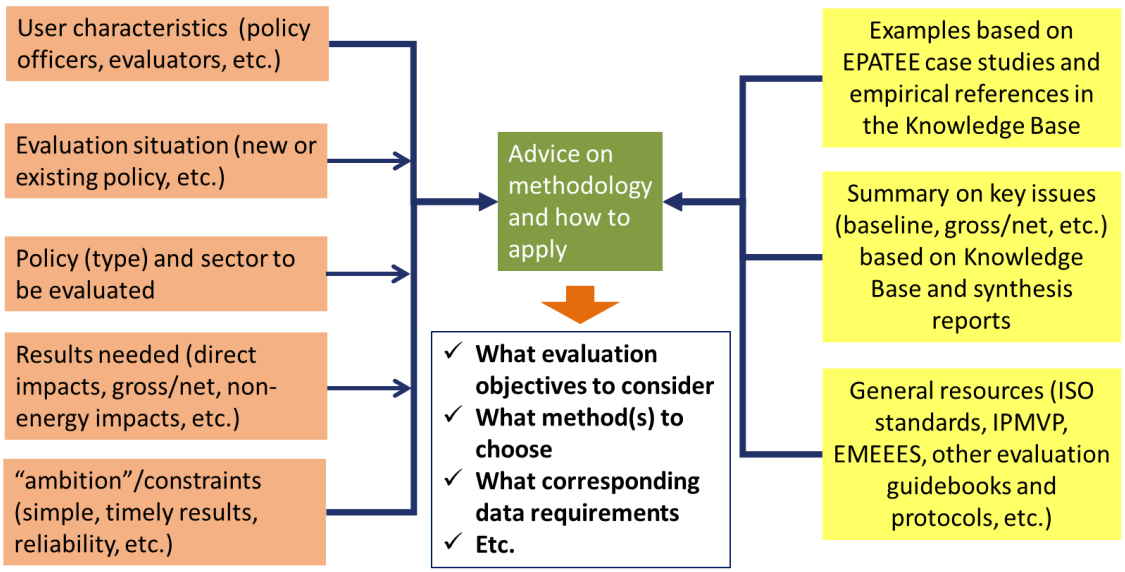


Figure 5. Outline of the EPATEE online tool

The different methods in the ex-post evaluation will be categorized in about 15 different approaches, these include, as an example:

- Different *Bottom-up methods*, for example:
  - Direct measurement of unitary energy savings
  - Unitary energy savings established on the basis of billing analysis
  - Deemed estimate of unitary energy savings
  - Detailed engineering estimate
- Different *Top-down methods*, for example:
  - Monitoring of energy consumption indicators
  - Econometric methods, simulation methods

Next to these methods, we also will make use of already accepted standards, that in some cases, can serve as a template for ex-post assessment of energy savings for specific regions or countries (cf. ISO 17742) or making use of developed protocols, as published by EVO (Efficiency Valuation Organization) in the form of IPMVP (international Performance Measurement and Verification Protocol).

We expect the online tool will be available as a test version in Q3 2018 and be operational on the EPATEE website by Q2 2019.

Looking at how evaluation can be effectively integrated into the policy cycle

Experience feedback collected in the case studies show that performing evaluation is not only about practical (e.g., data collection) or methodological (e.g., defining a baseline) issues. Organizational issues can be as important, and particularly when considering the planning and use of evaluation.

In parallel to the development of the online toolbox, the project will also aim at investigating the connections between evaluation and the policy cycle, and how to make evaluation an integral part of policy management. Examples of issues to be tackled are: what criteria to take into account when defining the evaluation objectives, how to ensure an early planning of the evaluation and in particular of the data

collection to reduce the evaluation costs, how to find the right timing for the evaluation, which stakeholders to involve in the evaluation process, etc.

Discussions at the workshop also emphasized the importance of evaluation costs (balancing costs and benefits of evaluation according to the evaluation objectives) and raised debates about external vs. internal evaluations, as well as about the definition of "independent evaluation".

Acknowledgements

The EPATEE project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement No 746265. The views expressed in this paper are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

The authors thank the reviewers for their useful comments to improve the paper.

References

Bini, Valentina, D'Ambrosio, Stefano, and Dario Di Santo, 2017. "Report on the interviews to the key stakeholders and the first EPATEE survey." Synthesis report, EPATEE project, November 2017. Available at: https://epatee.eu/main-results

Broc, Jean-Sébastien, Guermont, Catherine, Deconninck, Christian and Marie-Laure Nauleau, 2017. Impacts and cost-effectiveness of major energy efficiency policies for existing buildings: what do we exactly know and what can we learn? Proceeding of the ECEEE 2017 Summer Study, 1747-1758. Available at: https://proceedings.eceee.org/visabstrakt.php?event=7&doc=8-121-17

Broc, Jean-Sébastien, Bourges, Bernard, and Jérôme Adnot, 2007. Evaluation as a "Learning-by-Doing" Tool for the Implementation of Local Energy Efficiency Activities. Proceedings of the 2007 International Energy Program Evaluation Conference, 13-16 August 2007, Chicago. Available at: https://www.iepec.org/conf-docs/papers/2007PapersTOC/papers/2_1001_ab_512.pdf

EC, 2016. Proposal for a Directive of the European Parliament and of the Council amending Directive 2012/27/EU on energy efficiency. COM(2016) 761 final of the European Commission. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1485938766830&uri=CELEX:52016PC0761

EC, 2015. Assessment of the progress made by Member States towards the national energy efficiency targets for 2020 and towards the implementation of the Energy Efficiency Directive 2012/27/EU as required by Article 24 (3) of Energy Efficiency Directive 2012/27/EU. European Commission, COM(2015) 574 final. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:574:FIN

EC, 2013. Strengthening the foundations of Smart Regulation – improving evaluation. European Commission, COM (2013) 686 final. Available at: http://ec.europa.eu/smart-regulation/docs/com_2013_686_en.pdf

EEW, 2016. How to make Europe Number 1 in Energy Efficiency. Key results from the Energy Efficiency Watch project. April 2016. Available at: http://www.energy-efficiency-watch.org/index.php?id=257

EMEEES, 2009. "Evaluation and Monitoring for the EU Directive on Energy End-Use Efficiency and Energy Services". EU-funded project within the Intelligent Energy Europe programme. Results available at: http://www.evaluate-energy-savings.eu

EVO, 2012. International Performance Measurement and Verification Protocol – Concepts and Options for Determining Energy and Water Savings Volume 1. Efficiency Valuation Organization, January 2012. Available at: https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp

Haug, Constanze, Rayner, Tim, Jordan, Andrew, et al., 2010. Navigating the dilemmas of climate policy in Europe: evidence from policy evaluation studies. *Climatic Change* (2010) 101:427–445.

Hildén, Mikael, Jordan, Andrew and Tim Rayner, 2014. Climate policy innovation: developing an evaluation perspective. *Environmental Politics*, 23(5), 884-905.

Huitema, Dave, Jordan, Andrew, Massey, Eric, et al., 2011. The evaluation of climate policy: theory and emerging practice in Europe. *Policy Sciences*, 44 (2), 179–198.

Mundaca, Luis, and Lena Neij, 2010. A meta-analysis of bottom-up ex-ante energy efficiency policy evaluation studies. Proceedings of the 2010 International Energy Program Evaluation Conference, Paris. Available at: https://www.iepec.org/conf-docs/papers/2010PapersTOC/papers/033.pdf

NEEAPs, 2017. Second National Energy Efficiency Action Plans for the Energy Efficiency Directive. Reported by the EU Member States to the European Commission. Available at: https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive/national-energy-efficiency-action-plans

Schlomann, Barbara, and Wolfgang Eichhammer, 2011. Guidelines for the measure description in the MURE database. Version 3.1. September 2011. Available at: http://www.esd-ca.eu/Media/esdca/files/private/mure-guidelines

Wade, Joanne, and Nick Eyre, 2015. Energy efficiency evaluation: The evidence for real energy savings from energy efficiency programmes in the household sector. London: UK Energy Research Centre. Available at: http://www.ukerc.ac.uk/programmes/technology-and-policy-assessment/energy-efficiency-evaluation.html

Annex

Table 4. Data about energy savings and related calculation methods (from the EPATEE case studies).

Country	Name of the measure	reported energy savings	gross or net ?	Level 1 method	Level 2 method	Type of baseline	Adjustments, correction and other factors taken into account
Austria	Aid for environmental protection measures (UFI)	1.2 TWh/y from actions implemented over 2011-2013	gross	scaled savings	Method 5	“actual before”	No correction factors used Additionality criteria: performance > regulations; and payback time > 3 years (projects are additional, but savings are “gross” savings, as baseline = “actual before”)
Austria	City Energy Efficiency Programme (SEP) of Vienna	150 GWh/y (average rate of new annual final energy savings over 2006-2014)	additional	Deemed and scaled savings	Method 3 and method 5	“stock average”, “market average” or minimum energy performance requirements	Calculations according to the methods established in the frame of the Energy Services Directive (2006/32/EC)
Belgium (Wallonia)	Primes Energie (grants for energy renovation)	179 GWh/y in 2013 for actions implemented in 2013	gross	scaled savings	method 5	"stock average" for the characteristics of the building components (updated regularly)	use of normalised weather conditions and behaviours; no other adjustment factor applied; performance criteria on actions to ensure performance additionality
Croatia	Energy renovation programme for public sector buildings	Final annual energy savings in 2016: 0.177 PJ/y from actions implemented from 2014 to 2016; New annual final energy savings for actions installed in 2016: 0.053 PJ/y	gross	scaled savings	method 5	“actual before”	normalization of weather conditions, occupancy rates and operating hours

Country	Name of the measure	reported energy savings	gross or net ?	Level 1 method	Level 2 method	Type of baseline	Adjustments, correction and other factors taken into account
Croatia	Individual heat metering in multi-family buildings	Final annual energy savings of 0.119 PJ/y in 2016 from actions implemented from 2014 to 2016	gross	Deemed savings	Method 3	“actual before”(energy consumption before the installation of the heat cost allocators)	normalization of weather conditions no assessment of rebound effect, but it is noted that most buildings were over-heated before the installation of heat allocators
Denmark	EEO scheme	10961 TJ/y (about 3 TWh/y) for first-year final energy savings achieved from actions implemented in 2016	net	deemed or scaled savings	method 3 or method 5	"before" energy consumption (except for replacement of equipment where repair work cost > 25% of replacement cost, then baseline = market average or legal requirement)	Deemed savings are normalized (e.g., weather conditions, heating behaviours); Scaled savings are adjusted for changes in operation hours, production volumes, etc.; Conversion factors (for substitution between energy sources); Reduction factors (based on additionality assessments done in previous ex-post evaluations); Prioritisation factors (to favour some action types, e.g. actions with longer lifetime).
Finland	Energy Efficiency Agreement for Industries	11.1 TWh/y achieved in 2016 from actions implemented over 2008-2016 and still operating in 2016	gross	scaled savings	method 5	"before" energy consumption (or minimum energy performance standards when actions covered by EcoDesign)	Double counting with other policy measures is tracked;
Finland	Energy audits in municipalities	89 GWh/y in 2016 from actions implemented over 1995-2016	gross	scaled savings	Method 5	"before" energy consumption (or “actual before” when consumption has been metered)	Double counting with other policy measures is tracked;

Country	Name of the measure	reported energy savings	gross or net ?	Level 1 method	Level 2 method	Type of baseline	Adjustments, correction and other factors taken into account
Germany	Energy Efficiency Fund (data for the sub-measure "support for highly efficient cross-cutting technologies in SMEs")	525 GWh/year in 2016 for actions implemented over 2012-2016 (cumulated annual final energy savings)	Gross (but net results also evaluated)	deemed or scaled savings	method 3 or method 5	"before" energy consumption	Free-rider effects determined based on ex-post surveys. Double counting (interaction effects between the different sub-measures of the Fund)
Ireland	Better Energy Homes	Cumulative annual final energy savings: 994 GWh/year in 2016 (for actions implemented over 2009-2016)	gross	deemed savings	method 4	stock average (standard energy consumption per dwelling type)	Rebound effect (conservative values per type of dwelling, based on the comparison between modelled and metered energy consumption) Use of normalised weather conditions
Italy	White Certificates Scheme	about 2 Mtoe/y in 2016 (unit to be clarified)	additional	deemed, scaled or metered savings	methods 1, 3 or 4	highest energy performance from legal requirements, market average and before situation	Adjustments for industrial production, weather, plant or building usage, etc. (for scaled and metered savings). Baseline defined (and verified) to ensure savings are additional. Double counting (verifying certificates are not issued twice for the same action)
Lithuania	Renovation programme with EU funding	About 200 GWh/y (new final annual energy savings) from actions implemented in 2016	gross	scaled savings	method 5	actual before	Use of standardized heating behaviours and weather conditions; No adjustment (rebound effect, free-rider effect, etc.) is applied
Netherlands	Subsidy scheme for housing corportations in Amsterdam	About 0.9 Mm3 of gas saved/year from actions implemented over 2011-2014	gross		method 6	stock average	prebound effect (cases where, before implementing an energy efficiency action, end-users tend to consume less energy than estimated by engineering models)

Country	Name of the measure	reported energy savings	gross or net ?	Level 1 method	Level 2 method	Type of baseline	Adjustments, correction and other factors taken into account
Netherlands	Long-term agreement on energy efficiency for the non-ETS sector	60 PJ/y achieved in 2011 vs. 2005	gross	deemed, scaled or metered savings	methods 1, 2, 3, 4 or 5	actual before	Normalization for production volumes, weather conditions and possibly other factors depending on the type of action
Nordic countries	Nordsyn and the Effect project	18 GWh/year (avoided overconsumption for appliance sales in a typical year)	lost energy savings	deemed and metered savings	methods 1 and 4	minimum energy performance requirements	Normalised conditions of use (based on protocols used to test the appliances) Estimation of non-compliance rates
US	Weatherization Assistance Program	0.7 and 2.2 TWh/y from actions implemented respectively in 2008 and 2010	net	metered energy savings	Method 2	control group	Use of a quasi-experimental approach Normalization for weather conditions
UK	Supplier Obligations	about 10 TWh of final energy savings cumulated over the lifetime of actions implemented in 2015	gross	deemed savings	method 3	stock average, market average or minimum energy performance requirements	In-use factors accounting for rebound effects and performance gaps;
UK	Warm Front	8.0 TWh/year from actions installed over 2000-2010	gross	scaled savings	method 5	actual before	Calculations based on conventional energy consumption (normalised weather conditions and heating behaviours)

The tables below describe the typologies used to categorize the evaluation methods.

Table 5. Level 1 calculation methods.

Categories	**Explanations** (based on the Annex V of the EU Energy Efficiency Directive)
Deemed savings	"***deemed savings**, by reference to the results of previous independently monitored energy improvements in similar installations*"
Metered savings	"***metered savings**, whereby the savings from the installation of a measure, or package of measures, is determined by recording the actual reduction in energy use, taking due account of factors such as additionality, occupancy, production levels and the weather which may affect consumption*"

Categories	**Explanations** (based on the Annex V of the EU Energy Efficiency Directive)
Scaled savings	"***scaled savings**, whereby engineering estimates of savings are used (…) or where they are carried out on the basis of nationally established methodologies and benchmarks by qualified or accredited experts that are independent of the obligated, participating or entrusted parties involved*"
Surveyed savings	"***surveyed savings**, where consumers' response to advice, information campaigns, labelling or certification schemes, or smart metering is determined. This approach may only be used for savings resulting from changes in consumer behaviour. It may not be used for savings resulting from the installation of physical measures*"
Other	When not covered by one of the categories above (for example, in case of top-down methods, or bottom-up stock modelling ; see correspondences below)

Table 6. Level 2 calculation methods.

Bottom-up methods		***Link with level 1***
Method 1	Direct measurement of unitary energy savings (here, the unit usually is a participant)	"metered energy savings" or "surveyed savings"
Method 2	Unitary energy savings are established on the basis of billing analysis (unit = participant)	"metered energy savings" or "surveyed savings"
Method 3	Deemed estimate of unitary energy savings (the unit usually is a piece of equipment, but could sometimes be a participant if the end-use actions taken were rather uniform)	"deemed savings"
Method 4	Mixed deemed and ex-post estimate (e.g. unitary energy savings are based on equipment sales data, inspection of samples, monitoring of equipment purchased by participants) (the unit usually is a piece of equipment, but could sometimes be a participant if the end-use actions taken were rather uniform)	"deemed savings"
Method 5	Detailed engineering estimates (e.g., through calibrated simulation). This implies some more or less complex modelling of the individual unit (e.g. by calculating an energy balance of an individual building or an individual company in the dataset) (hence, the unit is normally a participant)	"scaled savings"
Mix methods		***Link with level 1***
Method 6	Stock modelling based on stock and market statistics, and surveys monitoring diffusion / uptake of enery-efficient solutions. This method will be a bottom-up method, if the surveys enable to identify, which end-use actions have been taken that change the energy consumption of the stock, and whether these end-use action were facilitated by EEI measures, and by which measures. Otherwise, this will be a top-down method	other
Method 7	Indicators of the share of specific equipment or practice in the market (diffusion indicators). Monitoring of these indicators will be a bottom-up method, if the change in indicator is entirely due to EEI measures (as is, e.g., the case for the installation of solar water heaters in many EU Member States). If this is not the case, and a regression analysis has to be performed to identify the energy savings due to EEI measures, this method will be a top-down method	other
Top-down methods		***Link with level 1***
Method 8	Monitoring of energy consumption indicators (either unit energy consumption for whole sectors or sub-sectors, or specific energy consumption indicators for specific end use equipment	other
Method 9	Econometric modelling (e.g., Input/Output analysis with price elasticities)	other
Combined bottom-up and top-down methods		*Link with level 1*
Method 10	Complex combinations of top-down and bottom-up methodologies in the form of integrated top-down and bottom-up methods	other