EPATEE

Evaluation into Practice: Lessons learnt from 23 evaluations of energy efficiency policies

Volume I: Main findings

Project Coordinator: Austrian Energy Agency – AEA

Work Package 3 Leader Organization: IEECP (Institute for a European Energy & Climate Policy)

August 2018

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement No 746265.

Authors

Jean-Sébastien Broc, IEECP

Vlasis Oikonomou, IEECP

Mia Dragovic, IEECP

With contributions by:

Gregor Thenius (AEA)

With thanks to:

All EPATEE partners for their hard work in preparing the case studies and their comments.

Dominique Osso (EDF R&D, France) for his comments.

Project coordination and editing provided by Austrian Energy Agency.

Manuscript completed in August, 2018

This document is available on: www.epatee.eu

Document title	Evaluation into Practice: Lessons learnt from 23 evaluations of energy efficiency policies – Volume I: Main findings
Work Package	WP3
Document Type	Public report
Date	31 August 2018
Document Status	Final version

Acknowledgments & Disclaimer

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 746265.

The EPATEE project

Several barriers limit energy efficiency policy evaluation. This results in a lack of quantitative data, and impedes evidence-based analysis required to distinguish effective from ineffective energy efficiency policies. EPATEE aims at tackling this problem by raising the capacity of policymakers and implementers. The project provides them both with tools and with 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, such as energy efficiency obligation schemes, regulations, financial incentives and voluntary agreements. Experience sharing is the core of the project. Lessons learnt from other EU initiatives and good practices in how to successfully evaluate the impact and cost-effectiveness of such energy efficiency policies will provide the basis for the development of guidelines and good practice evaluation tools. For further information please visit our website: www.epatee.eu

Executive Summary

This volume provides an overview of main findings about the different points analysed in the EPATEE case studies about evaluations of energy efficiency policies and programmes. The complete information from the case studies can be found in the volume II of this report (background report), ensuring the transparency about the findings presented in this volume.

The first part of this volume provides an overview of the case studies, including an analysis of their coverage in terms of policy instruments, sectors and countries. The case studies were not selected to provide a sample representative of the evaluation practices in Europe, but to cover the diversity of situations in view of experience sharing. Almost all main types of policy instruments and sectors are thus included in the set of case studies.

In terms of methods to evaluate energy savings, the case studies show that engineering calculations were the most frequently used for regular reviews or monitoring, whereas it was billing analysis for multi-year ex-post evaluations (apart from further analyses of the data from regular reviews). More generally, the choice of evaluation methods is strongly connected to the priority evaluation objectives (e.g., short-term reporting, verification of actual energy savings).

Data about energy savings mostly correspond to gross energy savings (i.e., using a baseline being the energy consumption before implementing actions and not applying any adjustment factors). In the majority of the 23 cases, factors used to assess net or additional energy savings (free-rider effects, spill-over effects or additionality) were indeed not evaluated or taken into account.

It is noticeable that no example of quantitative assessment of spill-over effects could be found among the 23 case studies. Likewise, information about analyses of market transformation effects was found in only 3 cases. This could therefore be a topic where further investigations could be needed.

In addition to the overview given in part 1, the systematic review of the case studies made it possible to draw lessons learnt for experience sharing, from a more qualitative and practical point of view:

- MESSAGE 01: Evaluation is not a burden, but an opportunity.
- MESSAGE 02: Evaluation priorities depend on who the primary audience is.
- MESSAGE 03: Evaluation helps increasing stakeholders' confidence in the schemes.

- MESSAGE 04: Monitoring and data collection are essential for making any evaluation possible.
- MESSAGE 05: Selecting the most relevant data to collect is a continuous process.
- MESSAGE 06: Regular review and in-depth ex-post evaluations are complementary.
- MESSAGE 07: The choice of evaluation methods depends on evaluation objectives but also on practical constraints.
- MESSAGE 08: Comparing different methods helps assessing the robustness of the results.
- MESSAGE 09: Evaluating net impacts is a challenge, but essential to assess efficiency of policies.
- MESSAGE 10: Good data is well-documented data.
- MESSAGE 11: Communication about evaluation results can be as important as doing the evaluation.

These messages are explained and illustrated with quotes and examples in part 2 to part 12 of this Volume.

Then, part 13 looks at the information collected beyond the impact evaluation of energy savings: indicators related to cost effectiveness and cost efficiency, impacts other than energy savings and other aspects than impacts. Feedback from stakeholders highlighted the interest and importance of expanding the scope of evaluation to consider other impacts then energy savings and to look at how the schemes work and why results are achieved or not.

Finally, three key debates raised along the interviews done for the case studies: independency of evaluation, if it is relevant to assess net energy savings and choosing between results based on estimates and results based on measured or metered data.

NOTE: the analysis presented in this report is not meant to be representative nor exhaustive about evaluation practices in Europe. It provides qualitative lessons learnt for experience sharing. It reflects what the stakeholders highlighted in the interviews done for the case studies, as well as what can be deducted from the review of the evaluation reports.

Table of Contents

1 | Introducing EPATEE case studies

1.1 Background and objectives of EPATEE case studies

The first motivation to do case studies about evaluations was that **experience sharing about evaluation practices in Europe is often limited** due to the lack of time for stakeholders to disseminate or document their 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; or the proceedings of IEPPEC, International Energy Policy & Programme Evaluation Conference). But most of the "regular" evaluation works remain in national language and are not always easy to find. Therefore, the actual evaluation practices of the stakeholders are not well known, and evaluation results are often disseminated without detailed explanations about their evaluation methods. This creates limitations for a correct understanding and use of evaluation results, as also noted by Haug et al. (2010) about climate policies, particularly for ex-post evaluations.

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

1.2 Methodology and content of the case studies

The selection of the case studies was **not meant to be exhaustive** in the sense of being representative for evaluation practices in Europe, **but to cover a diversity of situations** (policy instruments, sectors, countries, evaluation methods) and to provide interesting feedbacks in order to identify both, good practices and difficulties encountered.

A first set of 13 case studies was selected based on the knowledge of the EPATEE partners about evaluations in their country or in countries with a well-known experience with evaluation. Priority was given to evaluations of policies with an expected major impact and/or with a feedback that raised interest among other partners (which was assumed as showing the potential for experience sharing).

A second set of 10 case studies was later selected, taking into account suggestions from stakeholders and types of situations (policy instruments, sectors, countries, evaluation methods) that were not or little covered in the first set.

The analyses combined **two sources**. First, the **evaluation report(s)** and related information available (online or in paper documentation) were reviewed to fill in a template (see table below). Second, an **interview** was made with the evaluation customer and/or the lead evaluator to validate this summary and get a complementary experience feedback about the evaluation (reasons for the evaluation, lessons learnt about the policy, lessons learnt 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 was paid to explain the corresponding unit and scope (see Annex I). One frequent observation about data available online or in reports 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. Part of the added value of the EPATEE case studies is to make sure that the data are clear and explicit.

1.3 Overview of the case studies

The 23 case studies are available on the EPATEE website: https://epatee.eu/case-studies

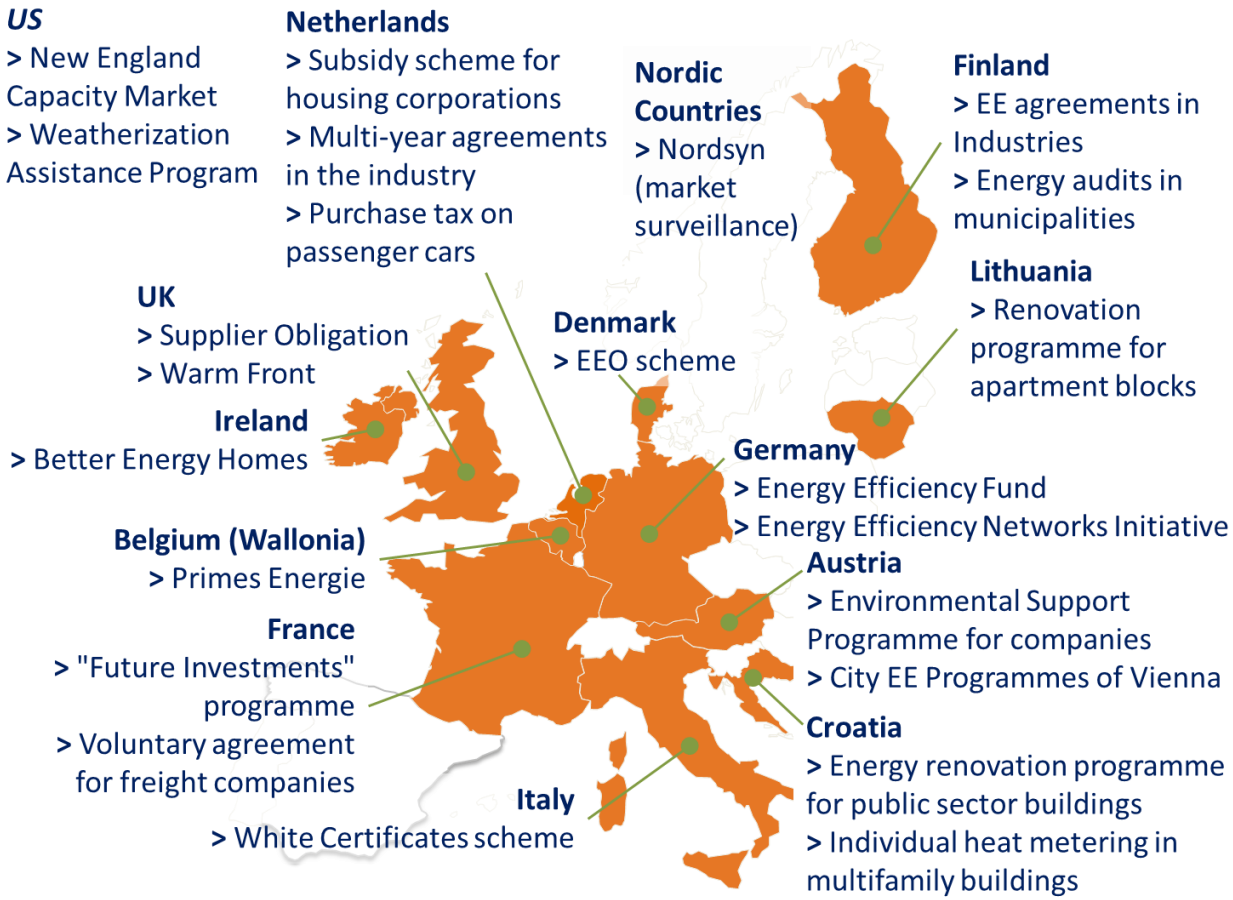


Figure 1. Map of the EPATEE case studies.

Table 2. List of EPATEE case studies.

Country	Sector	Type of instrument*	Name of the policy measure
Austria	Industry and services	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 metering in multifamily buildings
Denmark	Transversal	Market-Based	EEO** scheme
Finland	Industry and Services	Cooperative	Energy Efficiency Agreement for Industries
Finland	Services	Information/Education /Training	Voluntary energy audits for municipalities
France	Transport	Cooperative	Voluntary agreement for freight companies
France	Transversal	Financial	"Future Investments" programme
Germany	Industry and services	Cooperative	Energy Efficiency Networks Initiative
Germany	Transversal	Financial	Energy Efficiency Fund
Ireland	Residential	Financial	Better Energy Homes
Italy	Transversal	Market-Based	White Certificates Scheme
Lithuania	Residential	Financial	Renovation programmes with EU funding
Netherlands	Industry, agriculture and services	Cooperative	Multi-year agreements in the non-ETS sectors
Netherlands	Residential	Financial	Subsidy scheme for housing corporations in Amsterdam
Netherlands	Transport	Financial	Purchase tax on passenger cars
Nordic Countries	Residential	Legislative/normative	Nordsyn (market surveillance for the EcoDesign Directive)
UK	Residential	Market-Based	Supplier Obligations
UK	Residential	Financial	Warm Front
US	Energy sector	Market-Based	Auctions for capacity markets in New England
US	Residential	Financial	Weatherization Assistance Program

* typology taken from the MURE database, see Annex I and (Schlomann and Eichhammer, 2011).

** EEO: Energy Efficiency Obligation scheme

As shown in the figures below, all types of sector and instrument are covered by the set of case studies, except the type of instrument “legislative/information” (corresponding to mandatory audits, energy managers, building certificates or energy labelling).

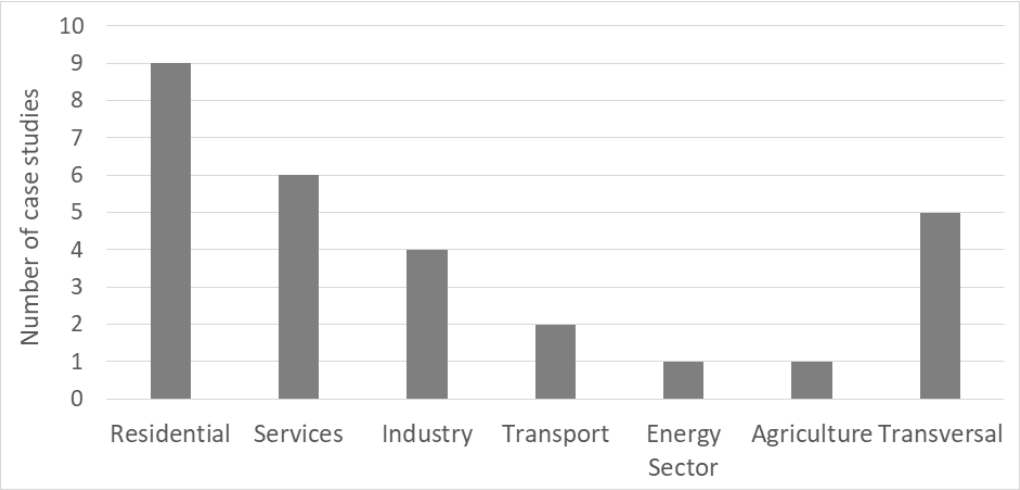


Figure 2. Distribution of the case studies per type of sector.

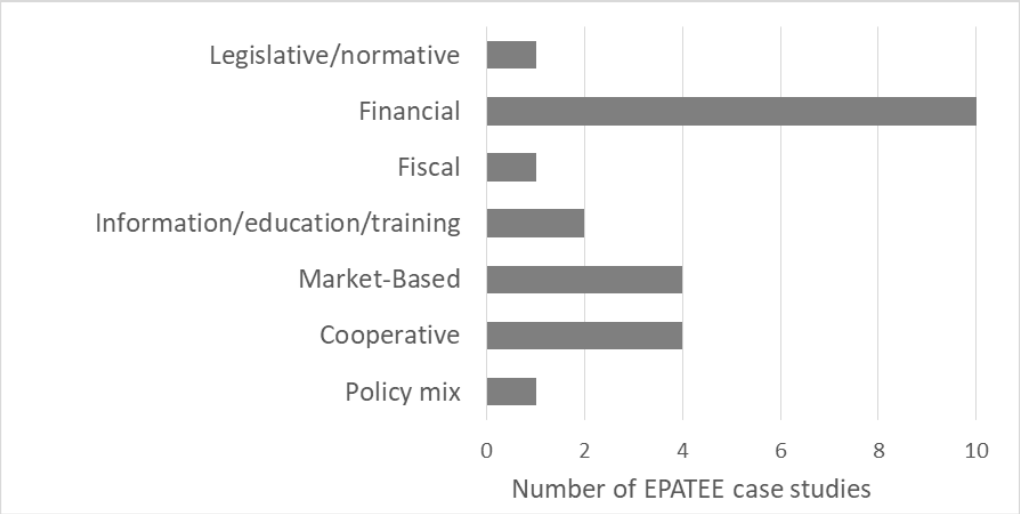


Figure 3. Distribution of the case studies per type of instrument.

Note: the typologies for sectors and instruments are presented in Annex I.

The highest numbers of case studies in terms of sectors and policy instruments are about the residential sector and financial measures respectively. This can be partly explained by the fact that these policies have also the highest frequencies in Member States' policy mix, as reported in the National Energy Efficiency Action Plans (NEEAPs) and analysed in the review published by the ODYSSEE-MURE project (Odyssee-Mure, 2015). About financial measures, this is also consistent with the feedback from the stakeholders' interviews: some of them indeed mentioned that evaluation efforts were prioritized on policies corresponding to high budget commitments (Bini et al., 2017). This is why it was easier to find cases about financial measures.

Some of the cases were selected because they dealt with policies representing a major contribution to achieving energy savings targets in Member States. This is another reason for the high frequency of financial measures, as well as market-based measures (particularly Energy Efficiency Obligation schemes) and cooperative measures (particularly voluntary agreements). This is also a reason for agriculture to be covered by a single case (that also covers industry and services), as most of the NEEAPs show low (or even no) contributions from measures in this sector.

The two cases on information measures were selected because they represent sub-types of instruments (voluntary energy audits, individual heat metering) that are directly related to articles of the EED (article 8 about energy audits, article 9(3) about heat metering). Then the low number of case related to information and behavioural measures can partly be explained because of the higher difficulty to prove energy savings from this type of measures, making that they are often considered in the policy mix as support measures. So their evaluation is often more focused on process evaluation, whereas the EPATEE project focuses on impact evaluation (about process or theory-based evaluation, see the case studies from the AID-EE project and Harmelink et al. 2008). The other reason is that examples of evaluations of behavioural measures can also be found in the resources of the IEA-DSM task 24. For recent developments in this field, see Frederiks et al. (2016).

Likewise, the fact that only one case is dealing with fiscal measures is because we didn't look for impact evaluations of energy or carbon tax, as a recent review on this topic was done for the European Commission (Europe Economics, 2016).

The low number of cases about transport, energy sector and legislative/normative measures (e.g., building codes) is due to difficulties to find examples about them.

About transport, one assumption could be that transport policies can be managed by a ministry different from the one in charge of reporting the NEEAP. Another assumption could be that energy savings from policy measures in transports can be more difficult to evaluate, as data can be more difficult to collect (e.g., fuel suppliers do not meter energy consumption for each end-user, but at best at the level of gas stations).

About the energy sector, policies on the supply-side of energy (generation-transmission-distribution) are covered by EU energy efficiency legislation only since the adoption of the Energy Efficiency Directive (see articles 14 and 15). Few policies for the energy sector have been reported so far in the NEEAPs. Therefore the single case study on this sector comes from the US. However, it should be noted that some of the transversal measures (EEO schemes in Denmark and Italy) allow actions in the energy sector under certain conditions.

About legislative measures, one assumption could be that these measures are very often subject to an evaluation before their adoption (or revision) (impact assessments or ex-ante evaluations). But they have been rarely evaluated ex-post (i.e. after being implemented). Whereas the focus of the EPATEE case studies was on ex-post evaluations.

1.4 Analysing the case studies

The primary focus of the analyses is about the evaluation of energy savings (impact evaluation). The two main questions investigated were:

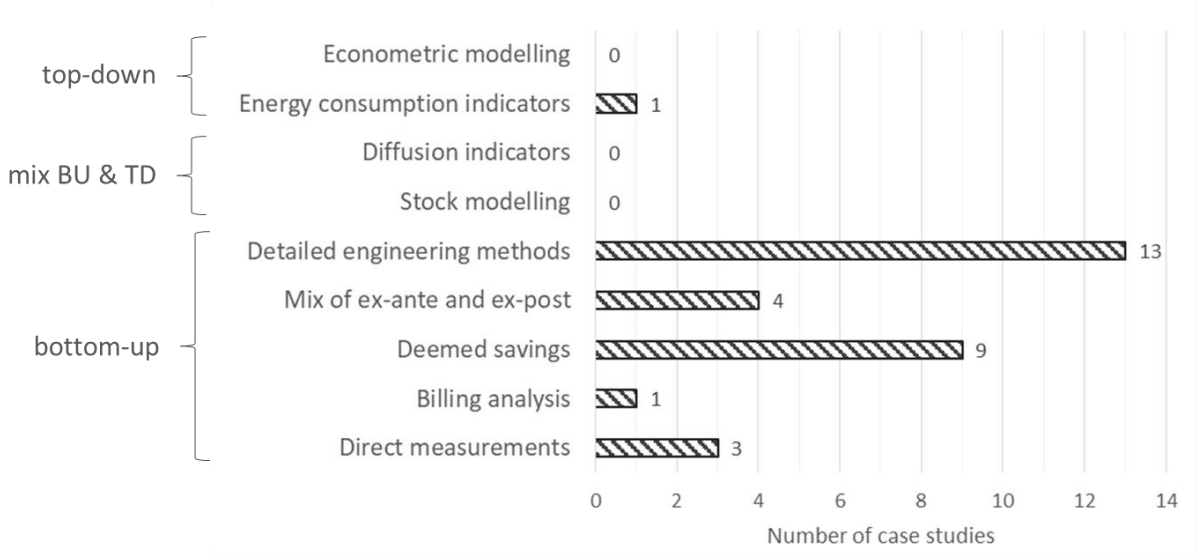
- **why (and what for) evaluations were decided and used**;
- **how they were performed**.

The structure of the template for the case studies was used to review all case studies in a systematic way, in order to identify the similarities and differences in the practices (based on the evaluation reports and related information) and in the feedback (based on the interviews with evaluation customers or evaluators).

The Volume I of this report presents the main findings about evaluation practices. Therefore it does not include all the detailed data and information collected in the case studies. They can be found in the Volume II of this report.

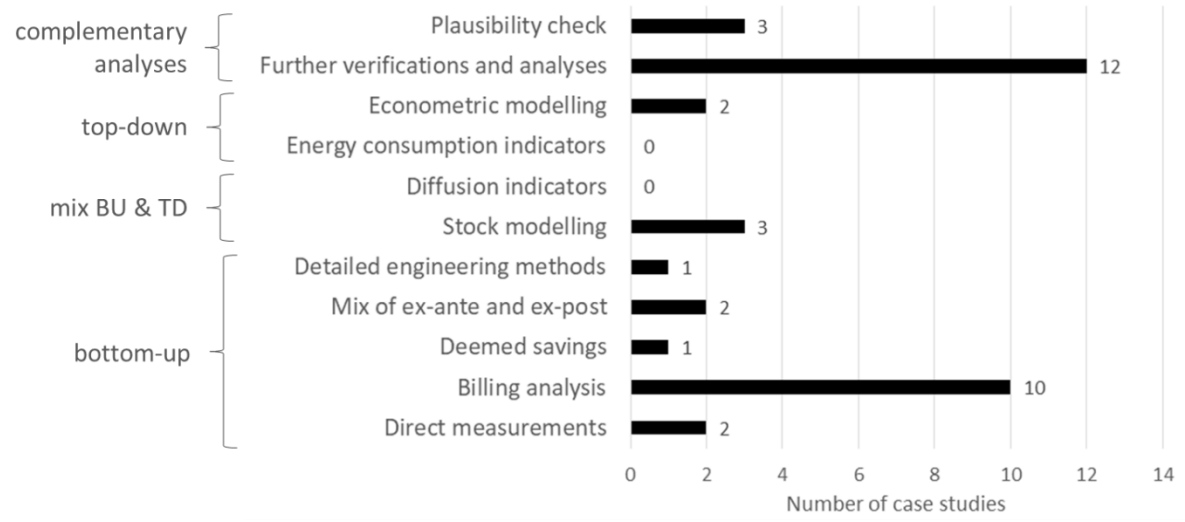
The figures below provide an overview of the criteria used to analyse the practices related to the evaluation of energy savings (primary focus of the case studies). For more details about the typologies used, see Annex I of this report.

About the types of **methods used to evaluate energy savings**, the data collected in the case studies led to distinguish the methods used for regular reviews or reporting (Figure 4) and the methods used for ex-post evaluations covering several years (or other ex-post studies) (Figure 5).



Note: one case can allow the use of different types of methods

Figure 4. Types of methods used for regular reviews or reporting.



*: complementary analyses based on the data from the regular reviews or monitoring (same methods as in Figure 4)

Note: one case can allow the use of different types of methods

Figure 5. Types of methods used for ex-post evaluations or studies.

Overall, only one case includes the use of a top-down method (energy consumption indicator), and only two cases includes the use of a method combining top-down and bottom-up approaches (stock modelling). This is because the EPATEE case studies were focused on ex-post evaluations of policies,

and thereby on bottom-up evaluations. Top-down evaluations are more frequently used to assess trends in energy consumption or energy efficiency indicators, or the achievement of national targets (e.g., targets of EED article 3).

The bottom-up methods correspond here to the methods used to evaluate energy savings at the level of a participant or project (unitary energy savings). In most cases, the number of participants or projects is directly collected through the monitoring system of the programme or policy (e.g., processing of applications for public aids, reporting from participants committed to voluntary agreements or from obligated parties to EEO schemes).

Figure 4 shows clearly that **engineering calculations** (i.e. deemed savings, mix of ex-ante and ex-post and detailed engineering calculations) are the methods the most frequently used to evaluate energy savings **for regular reviews**. This result is similar to the conclusions of Wade and Eyre (2015) and Labanca and Bertoldi (2016). This can be explained because methods based on measured (direct measurements) or metered (billing analysis) data need more time to provide results, as pointed in some case studies (see e.g., case on Warm Front). Whereas engineering calculations can be applied directly to data collected along the monitoring of the policy measure. Therefore, the need to report annually (or even more frequently) about energy savings often leads to choose to use engineering calculations.

At the opposite, Figure 5 shows that, besides complementary analyses of the data from regular reviews or monitoring, **billing analysis** is the method the most frequently used for **ex-post evaluations**. This can be explained because when ex-post evaluations include further efforts to evaluate energy savings, one of the main objectives is to assess actual energy savings (i.e. based on measured or metered data). The results from the billing analysis can then be compared to the results of the regular reviews, to improve the data or assumptions used in the engineering calculations. This is for example the approach that has been used in Ireland (Better Energy Homes), UK (Supplier Obligations) or in the US (Weatherization Assistance Program).

The two cases including **direct measurements** correspond to policy measures about market surveillance for appliances (Nordsyn in the Nordic countries) and purchase tax on new cars (the Netherlands). In both cases, energy savings are usually monitored based on standard energy consumption from energy labelling. The ex-post evaluations included a review of laboratory tests (appliances) or field measurements (new cars) to **investigate non-compliance rates and performance gaps respectively**.

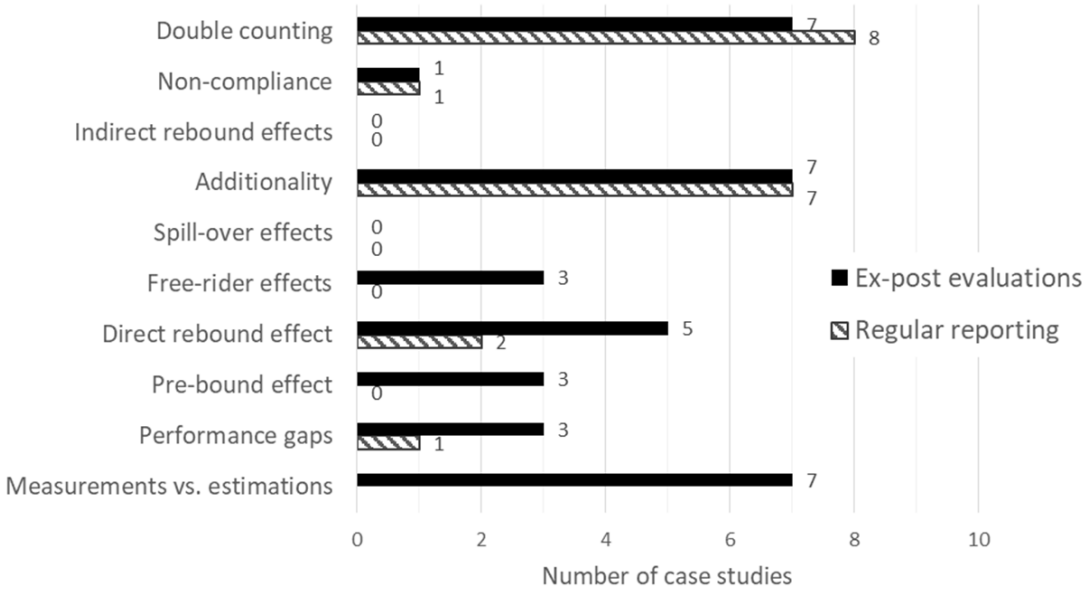
The distributions of types of methods do not necessarily reflect the practices of all EU countries and for all types of situations, as the sample of case studies is not meant to be statistically representative. However it confirms that **the choice of evaluation methods is strongly connected to the priority evaluation objectives** (e.g., short-term reporting, verification of actual energy savings).

Figure 6 shows that the types of correction or adjustment factors the most frequently taken into account are **double counting and additionality**, both usually aiming at ensuring that the energy savings accounted for the policy measure are additional to the results from other policy measures in place. For EEO schemes, double counting is tracked to avoid counting several times the same action that would be reported by different obligated parties or stakeholders applying for white certificates.

When these factors are taken into account in the regular reporting, they are most often included as well in the ex-post evaluations that might perform further verifications on these issues.

Factors that are used to correct for possible **differences between standard assumptions used in engineering calculations and actual conditions** (e.g., how the energy efficient products are installed and used) are more frequently taken into account in ex-post evaluations (see performance gaps, pre-bound and direct rebound effect). Indeed, when they are taken into account in regular reporting, it is

through default values usually based on previous ex-post studies (see for ex., Supplier Obligations in UK).



Note: one case can include calculations taking into account several factors.

Figure 6. Factors taken into account in the energy savings calculations.

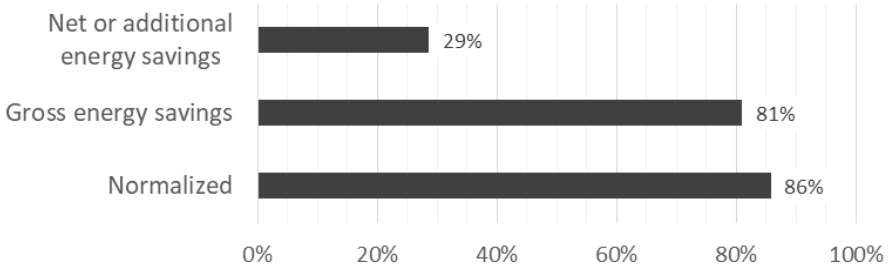
Figure 6 also shows that in the majority of the 23 cases, **factors used to assess net or additional energy savings** (free-rider effects, spill-over effects or additionality) are not taken into account. This is confirmed by Figure 7 that shows that in most cases, energy savings found in regular reporting are gross energy savings.

No case included any assessment of **indirect rebound effects**. This can be explained because these effects are usually evaluated at macro-economic level, using top-down approaches (vs. focus of the case studies on bottom-up evaluations).

One attempt to assess **spill-over effects** was found (Danish EEO scheme). However the evaluators concluded that the results from the survey could not be used to draw conclusions, and that further data collection would be needed. One assumption to explain the absence of assessment of spill-over effects is the difficulty to get access to market data (often very expensive).

Figure 7 below presents the types of energy savings that could be found in the regular reporting of the policy measures. It shows that in most cases, the results reported correspond to normalised gross energy savings. The high frequency of **normalised energy savings** can be explained by the high frequency of engineering calculations (for regular reporting). These calculations indeed often use assumptions about normalised conditions (e.g., normalised weather conditions, normalised heating behaviours).

The high frequency of **gross energy savings** and low frequency of **net energy savings** can be explained by the challenges faced when evaluating net energy savings that requires ex-post studies (e.g., ex-post surveys or billing analysis comparing participant and control groups) (see more details in part 10 |).



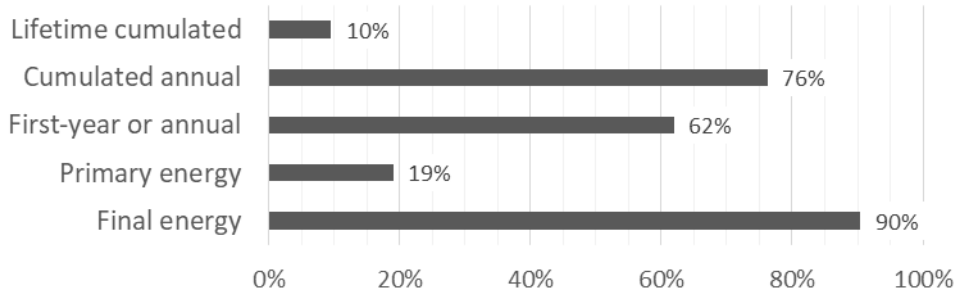
Notes:
- *one case can report energy savings in different units.*
- *results for 21 case studies (the cases on the French "Future Investments" programme and New England capacity market did not include data about energy savings).*

Figure 7. Type of energy savings (for regular reporting).

The low frequency of additional energy savings could be questioned, as some case studies show that it is feasible to define baselines enabling to calculate additional energy savings (see for ex., the Energy Efficiency Networks Initiative in Germany, or the white certificates scheme in Italy) or to include special factors to take into account additionality (see for ex., Danish EEO scheme).

In some cases, additionality criteria are used to ensure the additionality of the energy performance of the actions. Energy savings are then calculated against a baseline equivalent to the situation before implementing the actions (and without applying any adjustment factor), thereby reporting gross energy savings (see for ex., the two Austrian cases, Primes Energie in Belgium).

In terms of unit, the most common energy units used were multiples of GWh and PJ. As shown in Figure 8, energy savings are most often reported in annual terms, (either first-year or cumulated annual). Primary energy was used in few cases, mostly when there are national targets in terms of primary energy or when the policy also covers renewable energy actions (e.g., multi-year agreements in the Netherlands). In some cases results are directly reported in CO_2 savings (e.g., for the two cases on transports).



Notes:
- *one case can report energy savings in different units.*
- *results for 21 case studies (the cases on the French "Future Investments" programme and New England capacity market did not include data about energy savings).*

Figure 8. Units used to report energy savings.

In addition to the overview shown above, the systematic review of the case studies made it possible to draw lessons learnt for experience sharing, from a more qualitative point of view. These lessons are summarized in messages giving the titles of the next parts of this report.

The three first messages are about the strategic dimension of evaluation, based on stakeholders' feedback about the rationale and role of evaluation: why evaluation is done, what it is used for, who is involved and why evaluation can be important for stakeholders of energy efficiency schemes.

Then the messages correspond to lessons learnt along the evaluation process (parts 2 to 12):

- Organizing the data collection and monitoring → messages 04 and 05
- Planning and organizing the evaluation → message 06
- Choosing and comparing evaluation method(s) → messages 07, 08 and 09
- Reporting and communicating the evaluation → message 10 and 11

Then, part 13 looks at the information collected beyond the impact evaluation of energy savings: indicators related to cost effectiveness and cost efficiency, impacts other than energy savings and other aspects than impacts.

Finally, part 14 summarizes three key debates raised along the interviews done for the case studies: independency of evaluation, if it is relevant to assess net energy savings and choosing between results based on estimates and results based on measured or metered data.

NOTE: the analysis presented in this report is not meant to be representative nor exhaustive about evaluation practices in Europe. It provides qualitative lessons learnt for experience sharing. It reflects what the stakeholders highlighted in the interviews done for the case studies, as well as what can be deducted from the review of the evaluation reports.

2 | MESSAGE 01: Evaluation is not a burden, but an opportunity

> "These evaluations aim to investigate the satisfaction of the stakeholders (…), the impacts in terms of energy savings and the cost-effectiveness of the scheme (from a society point of view). The first point is important to know how the scheme can be improved in practice. The second point is important to know if the scheme meets its objectives. And the last point is to know if the energy distributors are using cost-effective approaches to deliver the energy savings."
>
> Quote from the case on the Energy Companies' Energy-Saving Efforts (Denmark)

This message is based on the review of why (and what for) evaluations were decided and used. Indeed, **none of the stakeholders interviewed for the case studies expressed a negative feedback** about doing evaluation. This is interesting to note, as the lack of interest of top management was mentioned as a possible important barrier to evaluation in the first EPATEE survey of stakeholders (see Bini et al., 2017).

The reasons and objectives to do evaluation were analysed according to the two dimensions classically used to characterize general evaluation objectives:

- the summative dimension, "**what are the results or impacts?**": assessing and reporting results, effectiveness and efficiency of the policies;
- the formative dimension, **"what can we learn or improve?"**: examining what works, what does not work, looking for improvements and questioning new ideas

Most evaluations analysed cover both dimensions to some extent, as illustrated in the quote above about the Danish case. **The main difference lies in the focus or priorities of the evaluation**: accountability (e.g., to the Ministry of Finance, the Parliament or the Court of Auditors), monitoring target achievements, assessing cost-effectiveness, getting a feedback about the satisfaction about the scheme, providing inputs to the redesign or improvement of the scheme, … (see more details in the section 2.1 of Volume II of this report).

Most of the evaluations analysed had multiple objectives. However, none of them aimed at covering all the evaluation criteria listed in the evaluation toolbox of the European Commission. Indeed, evaluation questions most often need to be prioritized. As mentioned in the feedback about the Environmental Support scheme in Austria for which all the evaluation objectives initially considered would have required a budget three times higher than the one available. Evaluation priorities can depend on its audience (see next part).

The case studies also gathered various testimonies and practical examples showing the added value of evaluation, as illustrated in the table below.

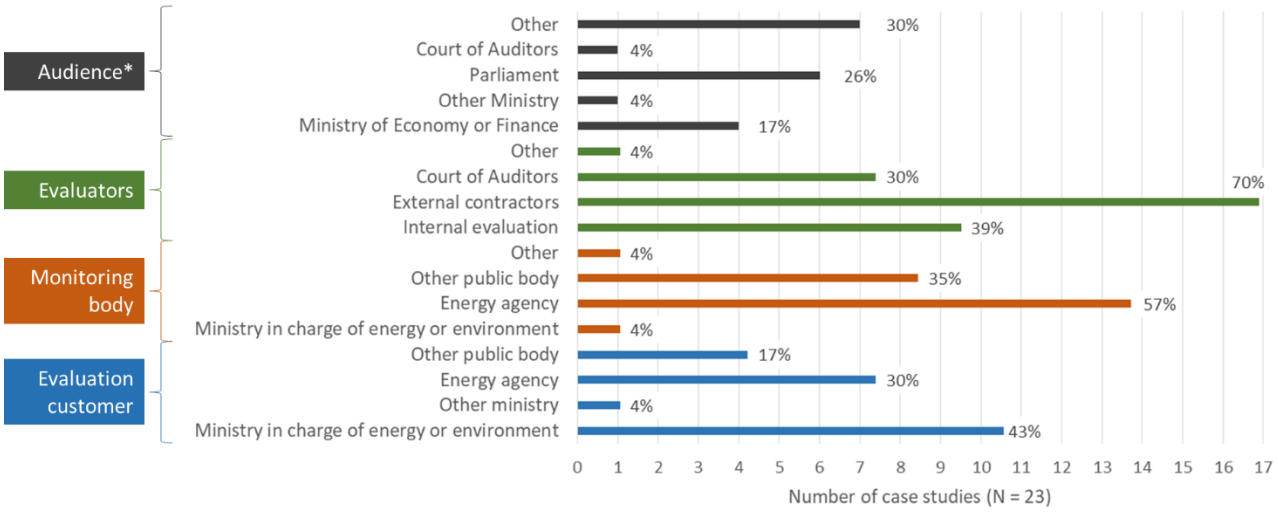
Table 3. Examples of use of evaluation results, conclusions or recommendations.

Examples of outputs/outcomes from the evaluation	Cases where these examples were mentioned
Political outputs	
Evidences/accountability for decision-making (particularly about funding)	Better Energy Homes (IE), EE Fund (DE), Environment Support Scheme (AT), Individual heat metering (CR), Voluntary energy audits (FI), White Certificates scheme (IT), WAP (US)
Reinforcing support from policymakers and other stakeholders	Better Energy Homes (IE), Voluntary agreements (FI), Voluntary energy audits (FI), Nordsyn, WAP (US)
Improving policy management	
Optimising the programme management	EE Programmes of Vienna (AT), Renovation programmes (LT), Supplier Obligation (UK)
New components added to increase scheme participation	Voluntary agreements (FI), Renovation programmes (LT), Supplier Obligation (UK)
Improving the application process	Primes Energie (BE), Environment Support Scheme (AT)
Improving monitoring and conditions for future evaluations	EE Programmes of Vienna (AT), EEO scheme (DK), Agreement for freight companies (FR), "Future Investments" programme (FR), Better Energy Homes (IE), Nordsyn, WAP (US)
Adapting the scheme and its rules	
Redesign of the incentives	Energy renovation of public sector buildings (CR), Individual heat metering (CR) Environment Support Scheme (AT), Renovation programmes (LT)
Improving data collection and verification processes	EEO scheme (UK), Environment Support Scheme (AT), Agreement for freight companies (FR), "Future Investments" programme (FR), Supplier Obligation (UK)
Updating the list of eligible actions	Primes Energie (BE), EEO scheme (DK)
Improved technical recommendations/requirements	Warm Front (England), Environment Support Scheme (AT), Voluntary energy audits (FI), EE Fund (DE), Multi-year agreements (NL), Warm Front (UK), WAP (US)
Better understanding of how the scheme works	
Reactivity of households to changes in the incentive design	Primes Energie (BE)
Detecting new trends and changes	Environment Support Scheme (AT)
Better understanding of interactions between policies	Voluntary energy audits (FI)
Better understanding of the reasons to participate (or not participate) to the scheme	Agreement for freight companies (FR), Renovation programmes (LT)
Understanding of interactions between policies	Voluntary energy audits (FI)
Understanding reasons of innovations success and failures	Agreement for freight companies (FR)
Understanding impacts and side-effects of the policy	Purchase tax on new cars (NL), Supplier Obligation (UK), Warm Front (UK), WAP (US)

3 | MESSAGE 02: Evaluation priorities depend on who the primary audience is

> "Choosing the most appropriate approach should consider the constraints (budget and time) and the evaluation objectives. What would be the most useful for the evaluation customers/recipients: knowing the impacts? or understanding the impacts?"
>
> Quote from the case on Warm Front (UK/England)

This message is based on the review of who was involved in the evaluations analysed, and of the general evaluation requirements. This shows the diversity in the organisation and role of evaluation. For more details, see also sections 3.1 and 4.1 of the Volume II of this report.



*: audience other than evaluation customers, monitoring body and evaluators
Note: one case can include several evaluations/evaluators and different actors in the audience. Only one case includes two different evaluation customers (for different evaluation studies, see individual heat metering in Croatia). Only one case includes several monitoring bodies that are all "other public bodies" (see Energy Efficiency Fund in Germany).

Figure 9. Who is involved in evaluations (and how).

Energy (or environment) ministries and public agencies are the most frequent **evaluation customers**. Other public bodies commissioning evaluations include energy authorities (or regulators) and local authorities (for local schemes). In some cases, evaluation has been done on the own initiative of an independent public body: Court of Auditors (e.g., Subsidy scheme for housing corporations in Amsterdam), or Environment Agency (see the case of purchase tax on new cars in the Netherlands).

Bodies in charge of the **monitoring** are most often energy agencies (57%) or other implementing bodies such as public banks. The scheme is directly monitored by a ministry in only one case (Primes

Energy in Belgium-Wallonia). The same for non-public body (see Warm Front that was monitored by an administrator contracted by the ministry).

About 70% of the cases include at least one **external evaluation**. Internal evaluation has been done in about 40% of the cases. Three cases include both, internal and external evaluations. The share of external evaluations is probably higher in the case studies than in the general practices in Member States, as we looked particularly for cases including published evaluation reports (which is less frequent in case of internal evaluations).

It can be noted that a review or evaluation by the **Court of Auditors** was mentioned in close to one third of the cases. These cases correspond to measures committing high budgets (either directly from public budget, or from obligated parties' costs in EEO schemes).

The information about the audience of the evaluation was not explicitly asked to the stakeholders interviewed. But several of them mentioned it along the interview. The picture above about the audience can therefore be incomplete.

Further analyses would thus be needed about the links between evaluation audience and evaluation priorities and scope. This will be done in the next phase of the project.

However the feedback collected from the interviews includes statement such as the one above that supports the assumption that evaluation questions could be prioritized according not only to the needs of the evaluation customers, but also to the perspective of the audience. For example, when the evaluation is reported to the Ministry of Economy or Finance, the evaluation has a focus on cost-effectiveness or related indicators. When the Court of Auditors is involved, questions related to value for money is often on the review or evaluation agenda.

> "Evaluations serve as a means of justification for actions taken. The Bundesrechnungshof (German Federal Court of Auditors) often uses evaluation reports for their judgement about the usage of public funding."
>
> Quote from the case on the Energy Efficiency Fund (Germany)

Who is the audience of the evaluation might also be a driver for the evaluation to cover specific aspects or impacts other than energy savings (see part 13 |). This will also be investigated further in the next phase of the project.

Unfortunately, the case studies rarely succeeded in collecting data about evaluation budgets or costs. But the feedback collected from the interviews pointed several times limitations in the evaluation scope or ambition due to the evaluation budget. This raises an assumption that the differences in the context of evaluation might not only lead to different evaluation objectives, but also different levels of resources dedicated to evaluation, and therefore different evaluation focus, expectations in terms of indicators and accuracy. This assumption could also be further investigated in the next phase of the project.

4 | MESSAGE 03: Evaluation helps increasing stakeholders' confidence in the schemes

> "One may have fear to do an ex-post impact evaluation, because it may show smaller results than based on the engineering estimates. However this increases the robustness of the results and therefore the confidence funders can have in them "
>
> Quote from the case on Better Energy Homes (Ireland)

Several interviewees raised the issues of **confidence, legitimacy or credibility** related to evaluation. Their experience is that stakeholders have confidence in the results if they trust the evaluators and their methods. Several testimonies also show that confidence in the evaluation results and conclusions turns into strengthening stakeholders' confidence in and support to the energy efficiency schemes. This can then translate into securing or increasing the funding for the scheme (as mentioned in the quote above from the Irish case) or the involvement of stakeholders such as participants to voluntary agreements (see Finnish case) or obligated parties of Energy Efficiency Obligation schemes (see Danish case).

Criteria and good practices to increase the credibility of or confidence in the evaluation are presented in the table next page. One of these criteria raises the issue of what "independent" can mean when speaking of **"independent" evaluation**, and why it is an important issue. It should be noted that the feedback about this issue included different points of view. This debate is summarized in part 14 |.

The topic of confidence in evaluation results can also be linked with discussions raised at the first European EPATEE workshop in Paris about the fact that decision makers or policy officers do not necessarily need results to be as accurate as scientific knowledge would allow (EPATEE, 2017). They need results to be robust enough in view of the decisions to take. This was also one of the key messages from the literature review made by Wade and Eyre (2015).

Another essential point is that stakeholders can consider results as reliable only if they are presented in a clear (i.e. transparent and understandable) way. This is related to message 10 about documentation of results.

For more details about the issue of credibility of and confidence in evaluation, see sections 3.2 and 3.3 of the Volume II of this report.

Table 4. Examples of criteria or good practices about credibility of and confidence in the evaluation results.

Examples	Cases where the examples are mentioned
Examples related to independency	
Evaluation commissioned to external/independent evaluators	Environmental Support Scheme (AT), EE Programmes of Vienna (AT), EE Network Initiative (DE)
Evaluation done or supervised by a service/department different from the one implementing the scheme	Primes Energie (BE), EE Programmes of Vienna (AT), Agreement for freight companies (FR), "Future Investments" programme (FR)
Possibility for an independent public body to do evaluations on its own initiative	Subsidy scheme for housing corporations (NL), Purchase tax on new cars (NL)
Examples related to including discussions along the evaluation process	
Steering group (including evaluation experts and/or stakeholders) involved in the evaluation process	Environmental Support Scheme (AT), EEO scheme (DK), WAP (US)
Including stakeholders in the discussions to prepare the evaluation (about objectives and methodologies)	EEO scheme (DK)
Regular exchanges between evaluators and evaluation customers (and possibly stakeholders) along the evaluation process	EEO scheme (DK), Environmental Support Scheme (AT)
Examples related to selecting the evaluators	
Skills and proven experience of the evaluators (clear criteria to select external evaluators)	Environmental Support Scheme (AT), Voluntary agreements (FI), EE Fund (DE)
Examples related to data quality and assumptions used in the energy savings calculations	
Clear procedure to verify the data submitted by participants	Environmental Support Scheme (AT), Voluntary agreements (FI), Voluntary energy audits (FI), White certificates (IT), Capacity Markets (US)
Training of stakeholders to ensure quality in data reported to the monitoring body	Voluntary agreements (FI), Voluntary energy audits (FI)
Examples related to approaches used to ensure the quality of evaluation results and conclusions	
Plausibility checks of the results	Environmental Support Scheme (AT), EE Programmes of Vienna (AT)
Comparison of different evaluation methods	"Future Investments" programme (FR)
Fostering the use of measured or metered data for the evaluation of actual energy savings	Better Energy Homes (IE), Energy renovation of public buildings (CR), Individual heat metering (CR), White certificates (IT), Renovation programmes (LT)
Use of conservative assumptions or correction factors to avoid over-estimations (of energy savings)	Nordsyn, Supplier Obligations (UK)
Examples related to providing favourable conditions for the evaluation	
Providing evaluators with all information needed so that conclusions reflect the reality of the scheme	Environmental Support Scheme (AT), Voluntary agreements (FI)
Providing good conditions for high-quality investigations (cf. evaluation budget and time)	Warm Front (UK), WAP (US)

5 | MESSAGE 04: Monitoring and data collection are essential for making any evaluation possible

> "Without robust monitoring data produced by our monitoring systems, the long-running scheme (since 1997) would not have continued this long, probably for just a few years."
>
> Quote from the case on Energy Efficiency Agreement for Industries (Finland)

Most of the feedback from stakeholders emphasised the importance of monitoring and data collection to provide the basis for evaluation. Experiences found in the case studies provide examples of difficulties encountered as well as good practices (see also sections 4.3 and 4.4 in Volume II).

Table 5. Examples of difficulties encountered with data collection or monitoring.

Examples of difficulties encountered	Cases where the examples are mentioned
Various reporting frameworks and timelines creating additional work	EE programmes of Vienna (AT), Primes Energie (BE)
Need to justify the resources (time and budget) needed to collect data / Budget and timelines limiting the possibility for data collection	Primes Energie (BE), EEO scheme (DK), Agreement for freight companies (FR)
Energy efficiency not always the priority objective of the scheme (some data needed to evaluate energy savings are not monitored)	Primes Energie (BE), Fiscal incentives for cars (NL)
Finding the right balance between procedures ensuring data quality and minimizing administrative burden	Primes Energie (BE), Multi-year agreements (NL)
Data providers might not trust the evaluators	Primes Energie (BE)
Technical issues to connect or match different databases / Technical issues with handling large data sets	Energy renovation of public buildings (CR), Primes Energie (BE), Better Energy Homes (IE), Subsidy scheme for housing corporations (NL)
Challenges to verify the situation before actions are implemented	EEO scheme (DK)
Data limitations with time series due to changes in data specifications over time, or to lack of disaggregated data	EEO scheme (DK)
Investment cost and time needed to develop web platforms	Voluntary agreements (FI), Voluntary energy audits (FI)
Changes in the projects or programmes along the evaluation process	"Future investments" programme (FR), WAP (US)
Time needed to observe the results of the actions	"Future investments" programme (FR), EE Network initiative (DE)
Conditions to access to billing data (e.g., privacy issues)	Better Energy Homes (IE), WAP (US)
Errors and frauds in data reported by stakeholders / data quality not ensured along the whole collection chain	EEO scheme (DK), White certificates (IT), Subsidy scheme for housing corporations (NL)
Finding the right contact to get the data (cf. large organisations, staff turnover)	Subsidy scheme for housing corporations (NL)

Table 6. Examples of good practices related to data collection and monitoring.

Examples of good practices	Cases where the examples are mentioned
Use of standardized data collection procedures	Environmental Support scheme (AT), EE programmes of Vienna (AT), Voluntary agreements (FI), Voluntary energy audits (FI), EEO scheme (DK), White certificates (IT)
Preparation work by the policy officer to ensure all information needed about the scheme will be available to the evaluators	Environmental Support scheme (AT), Voluntary agreements (FI)
Clear quality insurance processes (requirements, verifications)	Environmental Support scheme (AT) , White certificates (IT), Voluntary agreements (FI), Voluntary energy audits (FI)
One body in charge of coordinating data collection from various services or organisations	EE programmes of Vienna (AT), EE Fund (DE)
Clear definition of the indicators to be monitored (and related data requirements)	EE programmes of Vienna (AT), EE Fund (DE)
Use of web platforms to facilitate data collection / reporting	Environmental Support scheme (AT), Voluntary agreements (FI), Voluntary energy audits (FI)
Quality of the database(s) (clear structure, documentation, automatic or regular checks, compatibility with other databases)	Environmental Support scheme (AT), Primes Energie (BE), Energy renovation of public buildings (CR), Voluntary agreements (FI), Voluntary energy audits (FI), Supplier Obligations (UK)
Monitoring and verification procedures enforced by law	Energy renovation of public buildings (CR), Renovation programmes (LT), Supplier Obligations (UK)
Early planning/definition of data collection procedures and requirements	Voluntary agreements (FI), Voluntary energy audits (FI), Supplier Obligations (UK)
Training of stakeholders to ensure quality of data reported	Voluntary agreements (FI), Voluntary energy audits (FI)
Recognition of the commitment to the scheme is conditioned to meeting data requirements	Agreements for freight companies (FR), EE Network Initiative (DE)
Defining clearly who is responsible and subject to penalties (in case of problems with reported data)	White certificates (IT)
Suggestions that incentives should be conditioned to data requirements	Better Energy Homes (IE), Multi-year agreements (NL), Subsidy scheme for housing corporations (NL)
Cooperation between countries	Nordsyn
Taking into account the diversity in the complexity of the projects, by adapting data requirements to main types of projects	White certificates (IT), EEO scheme (DK), Capacity Markets (US)
Incentives that can cover the costs of strict data requirements	Capacity Markets (US)
Optimising the contacts with stakeholders (to avoid overburdening them)	WAP (US)

The review of how monitoring and data collection are organised reminds the no-brainer that it is essential to plan data collection when designing or adapting the policy measures.

> "The first point is a no-brainer, but always useful to remind: evaluation should be thought from the start, meaning when designing and starting each new policy. This is essential to organise the data collection and to ensure that the data needed for the evaluation will be available. This also helps to optimise costs for data collection."
>
> Quote from the case on Primes Energie (Belgium – Wallonia)

However, organising data collection upstream can be challenging. First because priority might be given to implementation, making that monitoring issues are dealt with later. Second because it is not always obvious to identify what data will be needed for further analysis (see next message about selecting the data to be collected).

More details about the linkage between monitoring and evaluation can also be found in the dedicated topical case study (Maric et al., 2018). More detailed analysis and guidelines about monitoring practices and tools can be found in the outputs of the multEE project: https://multee.eu/

6 | MESSAGE 05: Selecting the most relevant data to collect is a continuous process.

> "Data coming from the participating companies is the 'raw material' of evaluation. There is no possibility to make compromises in the quantity, quality or submission deadlines of this data. Ensuring timely submission of good quality data requires a lot of administrative work. This has involved a lot of discussions with the contact persons of participating companies and looking after."
>
> Quote from the case on Energy Efficiency Agreement for Industries (Finland)

As reminded in the quote above, an effective monitoring of a scheme requires collecting data from the participants. While the public authorities usually define **minimum data requirements**, it is also frequent to discuss the organisation and optimisation of the data collection with the stakeholders involved in the collection process.

Feedback on how the data to be collected are selected show that data collection procedures are **fine-tuned over time**. An efficient monitoring is indeed not only to plan the data collection early enough (or to put in place tools or platforms to collect data), but to select the data that represent the best cost/benefit ratio, in terms of data collection cost or burden vs. added value in terms of better reliability of the results and/or understanding of the policy and its impacts.

Feedback thus also shows that involvement of stakeholders in data collection is higher when they see what the data are used for, and that they get an added value from this process. **Identifying what data are useful and what data are not really needed** is critical to find the right balance between ensuring relevant and reliable monitoring and evaluation outputs, and costs of the data collection procedures.

Particular attention is indeed often paid to avoid creating a too high administrative burden that would reduce the participation to the scheme. At the same time, some of the interviewees also argued that when public money is given to participants, it is **legitimate to ask them to provide some minimum information** to make possible to monitor the results (see Belgian, Finnish and Irish cases for example). The same can apply when an obligated party provides participants with financial or other supports. Putting in place this kind of conditions should also take into account **privacy and data protection regulations**.

Moreover, data requirements are not only related to the need for evaluating energy savings. They come most often from needs related to quality and verification: verifying eligibility of the participants, performance criteria of the actions, skills of the installers, etc. One challenge might thus be to organise synergies between the different data needs, and to avoid separate data collections that impede data to be matched at the end.

Table 7. Examples of practices to fine-tune the data collection requirements or process.

Examples of practices to fine-tune the data collection requirements or process	Cases where the examples are mentioned
Regular analysis to detect changes and needs to adapt data requirements and collection procedures / collecting feedback from the stakeholders (e.g., open consultation, satisfaction survey)	Environmental Support scheme (AT), Voluntary agreements (FI), White certificates (IT), Supplier Obligations (UK)
Monitoring procedures reviewed during the external evaluations (including recommendations on missing data or data that are not used)	Environmental Support scheme (AT), EE programmes of Vienna (AT), EEO scheme (DK), Voluntary agreements (FI), "Future investments" programme (FR), Energie Primes (BE), WAP (US)
Asking data that can be easily verified (e.g., invoices)	Primes Energie (BE)
Efforts made to reduce the time needed to process the data submitted by participants/stakeholders	Environmental Support scheme (AT), Primes Energie (BE)
Organising exchanges with the different services/bodies involved in data collection and reporting to ensure coordination and all constraints and needs are taken into account	EE programmes of Vienna (AT)
Use of standardised templates or spreadsheets to harmonize data reporting among different services or bodies	EE programmes of Vienna (AT), EE Fund (DE)
Work on connecting different databases to have more data directly available	Primes Energie (BE), Energy renovation of public buildings (CR), Supplier Obligations (UK)
Analysis about the needs for controls or ex-post verifications	EEO scheme (DK), White certificates (IT)
Increasing the resources dedicated to processing/verifying the data when energy savings targets increase	EEO scheme (DK)
Providing a feedback to stakeholders so that they can see the added value of the monitoring process	Voluntary agreements (FI), Voluntary energy audits (FI)
Surveying stakeholders/participants about their interest in more feedback/benchmarking (to decide whether collecting more data)	EE Network initiative (DE)
Identifying the main sources of uncertainties and what data could help reduce them	Better Energy Homes (IE), Subsidy scheme for housing corporations (NL), Warm Front (UK)
Ensuring that the monitoring outputs are relevant to programme administrators	Renovation programmes (LT)
Objective of minimizing the administrative burden for the stakeholders	Multi-year agreements (NL)
Regular information of the stakeholders about the updates in the data collection procedures	Voluntary agreements (FI), Voluntary energy audits (FI), Multi-year agreements (NL)
Organising a framework that facilitate data contributions from different types of stakeholders	WAP (US)

7 | MESSAGE 06: Regular review and in-depth ex-post evaluations are complementary

> "The ex-post evaluations are used to complement the monitoring of the scheme when preparing a revision of the agreement for the scheme."
>
> Quote from the case on Energy Companies' Energy-Saving Efforts (Denmark)

The analyses done in the case studies led to distinguish two main practices related to monitoring and ex-post evaluation:

1. Regular reviews, usually done annually and based on the verification and compilation of data from on-going monitoring and/or annual reports prepared by participants (voluntary agreements) or obligated parties (energy efficiency obligation schemes).
2. In-depth ex-post evaluations, usually covering a multi-year period, and including further data collection (e.g., surveys, interviews) and analysis.

The case studies show a broad consensus on the need to implement regular reviews due to reporting needs, but also to get a quick feedback loop to detect potential problems.

> "If there are problems, we need to know where those are. It is another question if we can interfere, but we must know and understand the situation."
>
> Quote from the case on Voluntary audits for municipalities (Finland)

The need and periodicity or timing for in-depth ex-post evaluations can vary. 17 of the 23 case studies include at least a dedicated and official ex-post evaluation. In 6 cases, this was done only once. In 4 cases, ex-post evaluations are done upon request. For the remaining 7 cases, ex-post evaluations are organised on a regular basis, mostly at the end of each period of the scheme or due to reporting requirements (periodicity mostly included between 3 to 5 years). The two German cases include on-going evaluations. And for the Supplier Obligations in UK, ex-post evaluations are done at mid-term of the periods, to let time to take into account conclusions for the next period.

It should be noted that when selecting the case studies, a preference was given to cases including in-depth ex-post evaluations. Therefore there might be a bias in this sample in terms of frequency of in-depth ex-post evaluations.

Some of the interviewees pointed the difficulty to organise ex-post evaluations in a relevant timing, particularly because more time might be needed to observe the impacts of the scheme, whereas

consultation with stakeholders often requires to get conclusions and recommendations early enough when preparing the next period of a scheme.

> "The Supplier Obligation is continuous and there is typically not enough time to carry out an evaluation at the end and apply the learnings in the next phase. We therefore conduct post implementation reviews which happen mid-term during an ongoing phase of the Supplier Obligation. There is also continuous tracking of impacts of scheme in terms of actions and costs - this reduces the need for substantial evaluations at the end of the scheme."
>
> Quote from the case on the Supplier Obligations (United Kingdom)

Table 8. Examples of complementarity between annual reviews and multi-year evaluations.

Examples of complementarity between annual reviews and multi-year evaluations	Cases where the examples are mentioned
Annual reviews used to get a regular feedback look and monitor achievements. Ex-post evaluations used to analyse further the results and investigate more in details specific issues.	Environmental Support scheme (AT), EE programmes of Vienna (AT), EEO scheme (DK), EE Fund (DE)
Annual reviews/monitoring provide the basic data or starting point for ex-post evaluations.	Environmental Support scheme (AT), EE programmes of Vienna (AT), Primes Energie (BE), EEO scheme (DK), Agreement for freight companies (FR), EE Network Initiative (DE), EE Fund (DE), Purchase tax on new cars (NL), Supplier Obligations (UK)
Ex-post evaluations or studies used to assess actual energy savings (based on metered data), to verify (and possibly update) energy savings based on engineering calculations.	Individual heat metering (CR), Better Energy Homes (IE), Subsidy scheme for housing corporations, WAP (US)
Ex-post evaluations made to review the monitoring and verification procedures	Environmental Support scheme (AT), EE programmes of Vienna (AT), EEO scheme (DK), WAP (US)
On-going monitoring used to review data from each project. Ex-post evaluation used to assess the overall impact of the scheme.	"Future investments" programme, EE Fund (DE)
Annual reviews/monitoring focused on energy savings. Ex-post evaluations bringing complements to assess cost-effectiveness.	EEO scheme (DK), Better Energy Homes (IE), Agreement for freight companies (FR), Multi-year agreements (NL), Nordsyn
Ex-post evaluations or studies used to complement monitoring with qualitative analysis	Primes Energie (BE), Better Energy Homes (IE), Multi-year agreements (NL)

8 | MESSAGE 07: The choice of evaluation methods depends on evaluation objectives but also on practical constraints

> "Our experience is that when preparing a tender for an evaluation, the specifications for the evaluation should be focused on defining clear evaluation questions. The choice of the evaluation methods to answer these questions should be up to the bidders. This makes possible to compare offers with different methodologies."
>
> Quote from the case on the Energy Companies' Energy-Saving Efforts (Denmark)

> "While engineering estimates are useful to monitor the results on an on-going basis, I strongly recommend going beyond engineering estimates. One may have fear to do an ex-post impact evaluation, because it may show smaller results than based on the engineering estimates. However this increases the robustness of the results and therefore the confidence funders can have in them."
>
> Quote from the case on Better Energy Homes (Ireland)

The feedback collected in the case studies about the use of engineering methods and statistical methods shows that there is no "perfect" evaluation method to assess energy savings. All methods have advantages and limitations.

Table 9. Advantages of engineering and statistical methods and links with evaluation objectives or needs.

Type of method	Advantages	Usually chosen when the evaluation objectives or needs are…
Engineering methods	✓ Might enable to automatize energy savings calculations ✓ Estimations directly linked to the energy efficiency improvement	✓ On-going monitoring of the results ✓ Providing visibility to the stakeholders and participants ✓ Focus on the technical performance of the actions ✓ Evaluation of gross energy savings ✓ Possibility to take into account additionality criteria defined beforehand
Statistical methods	✓ Based on actual energy consumption (energy bills or other metered data) ✓ Might enable comparison between participants and control group	✓ In-depth ex-post evaluation ✓ Taking into account behaviours of end-users ✓ Evaluation of net energy savings

Table 10. Feedbacks about the use of engineering methods.

Difficulties [-] or other lessons learnt [+]	Cases where it is mentioned
[+] Enable on-going monitoring of the results	Better Energy Homes (IE)
[+] Possibility to use simple methods making the monitoring system easy to implement	Multi-year agreement (NL)
[+] Possibility to use detailed engineering calculations (scaled savings) to assess specific projects (particularly for the industry sector)	Environmental Support scheme (AT), White certificates scheme (IT)
[+] Possibility to improve data quality through training of stakeholders	Voluntary agreements (FI), Voluntary energy audits (FI)
[+] Possibility to improve reliability of modelling by using calibration on actual energy consumption	WAP (US)
[-] Uncertainties due to the assumptions used in the calculations (e.g. about behaviours)	EE programmes of Vienna (AT), EE Fund (DE), Renovation programmes (LT), Subsidy scheme for housing corporations (NL), Warm Front (UK)
[-] Difficulties to verify the baseline situation or update baselines	Primes Energie (BE), EEO scheme (DK)
[-] The higher the targets, the higher the number of calculations to review/verify, the more difficult to make systematic verifications	EEO scheme (DK), White certificates (IT)
[-] Estimated savings not always possible to verify due to difficulties to make measurements in practice or to the costs it would induce	Voluntary agreements (FI), Voluntary energy audits (FI)
[-] Cannot evaluate all types of actions (particularly behavioural actions)	EE programmes of Vienna (AT)

Table 11. Feedbacks about the use of statistical methods.

Difficulties [-] or other lessons learnt [+]	Cases where it is mentioned
[+] Providing results based on data of actual energy consumption	Better Energy Homes (IE), Subsidy scheme for housing corporations (NL), Individual heat metering (CR)
[+] Empirical verification needed to ascertain the estimations represent a small budget compared to the whole budget of the scheme	Better Energy Homes (IE)
[+] Roll out of smart meters could increase the availability of metered data	EEO scheme (DK)
[+] Possibility to use methods for sensitivity analysis when extrapolating results from a sample to the whole population	Individual heat metering (CR)
[+]/[-] Top-down analysis provides a better understanding of factors influencing trends in energy consumption, but it does not show what the contribution of the programme is.	EE programmes of Vienna (AT)
[-] Difficulties to get market data for market data analysis (too costly, data not detailed enough)	EEO scheme (DK)
[-] Difficulties to collect billing data for billing analysis (need to contact participants afterwards for agreement, willingness of utilities/availability of utility staff)	Better Energy Homes (IE), WAP (US)
[-] Difficulties to get disaggregated-enough data and consistent long-time series to perform econometric analyses	EEO scheme (DK), Warm Front (UK)
[-] Difficulties or even impossibility to define a relevant control or comparison group	EEO scheme (DK)
[-] Issue of representativeness of the samples/data	Agreement for freight companies (FR), Warm Front (UK)
[-] Costs (or practical issues) to get data from large enough samples to get results statistically robust/significant	EEO scheme (DK), “Future investments” programme (FR)
[-] Difficulties to control for all external factors (data needed to take them into account not available or too costly or complex to collect)	“Future investments” programme (FR), Multi-year agreements (NL), Warm Front (UK)
[-] Technical issues of handling large set of data (e.g., ensuring data protection)	Better Energy Homes (IE)
[-] Difficulties to combine results obtained using different sampling methods	Nordsyn
[-] Time-lag in getting results (data available only afterwards, need to get several years of data for results to be robust, time needed to collect and process data)	Supplier Obligations (UK), Warm Front (UK)

Ideally, the evaluation method(s) should be chosen according to the evaluation objectives (see Table 9 above). But practical constraints often also influence this choice: data availability, time and budget available for the evaluation, fields of expertise of the evaluators, requirements to use a particular methodology (e.g. national evaluation framework, guidelines of the European Commission), etc.

> "The main constraints were – as often is the case – related to time and budget limits. The main challenge in the 2008 and 2012 evaluations was to deliver sound results fast enough, with a very tight schedule. This implies finding ways to prioritize and select what work to do – in particular, in relation to the data collection."
>
> Quote from the case on the Energy Companies' Energy-Saving Efforts (Denmark)

Guidance about selecting evaluation method(s) depending on the context and objectives of the evaluation is one of the key subjects of the next phase of the EPATEE project.

9 | MESSAGE 08: Comparing different methods helps assessing the robustness of the results

> "In terms of evaluation methods, it is essential for us to use triangulation, i.e. to compare results obtained with different methods and/or from successive evaluations. This provides a stronger basis for our conclusions."
>
> Quote from the case on the Energy Companies' Energy-Saving Efforts (Denmark)

> "Empirical verifications represent a small budget compared to the whole budget of the scheme. Our experience with the ex-post impact evaluation is that it is really worth it."
>
> Quote from the case on Better Energy Homes (Ireland)

Several stakeholders interviewed emphasised the importance to compare results obtained with different evaluation methods or to perform plausibility checks. It is indeed often difficult to make a quantitative assessment of the uncertainties of results from using a single method.

Table 12. Examples of use or comparison of different methods.

Examples of use or comparison of different methods	Cases where these examples are mentioned
Plausibility check of the overall results (comparison with trends in energy consumption, and/or comparison with previous periods)	Environmental support scheme (AT), EE programmes of Vienna (AT), Primes Energie (BE)
Comparison of surveys and econometric analysis to assess additionality	EEO scheme (DK)
Comparison of different statistical methods	"Future investments" programme
Comparison of engineering calculations and billing data/analysis	Better Energy Homes (IE), Renovation programmes (LT), Subsidy scheme for housing corporations (NL), Supplier Obligation (UK), Warm Front (UK)
Comparison of monitoring of energy efficiency indicators (top-down approach) and monitoring based on engineering estimates at project level	Multi-year agreement (NL)
Comparison of standardised laboratory tests and field measurements	Purchase tax on new cars (NL)
Comparison of different methods to normalised energy consumption for weather conditions	WAP (US)

Table 13. Sources of uncertainties identified in the EPATEE case studies.

Sources of uncertainties identified in the EPATEE case studies	Cases where these types of uncertainties are mentioned
Errors in the data reported by participants of stakeholders	Environmental Support scheme (AT), Primes Energie (BE), EEO scheme (DK), White certificates (IT), Voluntary agreements (FI), Voluntary energy audits (FI), Multi-year agreements (NL), Supplier Obligations (UK)
Assumptions used for engineering calculations (about behaviours, possible performance gaps)	EE programmes of Vienna (AT), Primes Energie (BE), Energy renovation of public buildings (CR), EEO scheme (DK), Better Energy Homes (IE), Renovation programmes (LT), Supplier Obligations (UK), Warm Front (UK)
Differences between actual characteristics and reference values used (about baseline conditions, and/or energy performance of the actions)	Primes Energie (BE), Energy renovation of public buildings (CR)
Use of default values	Individual heat metering (CR), EEO scheme (DK), EE Fund (DE), Subsidy scheme for housing corporations (NL)
Statistical uncertainties due extrapolation from samples	Individual heat metering (CR)
Bias in the answers from surveys	EEO scheme (DK), EE Fund (DE)
Differences between actual performance in real conditions of use and standard performance assessed with normalised tests (e.g. manufacturers' data)	Voluntary agreements (FI), Voluntary energy audits (FI), EE Fund (DE), Purchase tax on new cars (NL)
Differences in the project as initially reported and the project finally implemented	"Future Investments" programme (FR)
Differences in calculation methods used by energy auditors (for complex actions)	EE Fund (DE)
Bias in sampling and/or matching methods	Better Energy Homes (IE), White certificates (IT), Nordsyn
Complexity of industrial processes (for very particular but large projects)	White certificates (IT)
Data limitations (sample size, time series)	EEO scheme (DK), Warm Front (UK)

About uncertainties related to data quality and monitoring, see part 5 | for examples of good practices to tackle them.

10 | MESSAGE 09: Evaluating net impacts is a challenge, but essential to assess efficiency of policies

> "The mid-term evaluation has also confirmed the difficulty in operationally implementing econometric methods and the difficulty in obtaining robust figures. However, it has helped understanding the causal processes leading (or not) to technological and/or commercial successes."
>
> Quote from the case on the "Investments for the Future" programme (France)

The case studies show a diversity in the way to define or handle the concept of net or additional impacts, as well as in the methods used. Difficulties were also often reported, sometimes impeding the use of the method initially planned. Some interviewees mentioned that there was not attempt to evaluate net energy savings because they expected uncertainties to be too high for the results to be relevant. This confirms that evaluating net impacts is a challenge. Stakeholders who tackled this challenge also reported that this is essential to assess and better understand the cost-effectiveness (or efficiency) of the policies.

Table 14. Feedbacks about the methods used to assess net or additional energy savings or impacts.

Feedbacks about the methods used to assess net or additional energy savings or impacts	Cases where these feedbacks are mentioned
Choice of an alternative approach: additionality criteria used to ensure the additionality of the energy performance of the projects/actions	Environmental support scheme (AT), Primes Energie (BE), EE Networks Initiative (DE)
Calculations providing additional energy savings, by defining baselines that takes into account additionality (minimum energy performance standards or market average) or applying special additionality factors	EE programmes of Vienna (AT), White certificates (IT), EEO scheme (DK)
Debates about the reliability of results from surveys	EEO scheme (DK)
Econometric analyses at sectoral level	EEO scheme (DK)
Use of quasi-experimental method(s) comparing samples of participants and non-participants	EEO scheme (DK), "Future investments" programme (FR), Better Energy Homes (IE)
Difficulties (or even impossibility) of defining or assessing a counterfactual	Multi-year agreement (NL), Subsidy scheme for housing corporations (NL), Supplier Obligations (UK)
Comparison with trends in other EU countries	Purchase tax on new cars (NL)
Causality assumed not to be an issue as the programme is targeted to low income households	Warm Front (UK), WAP (US)

For more details about methods that can be used to assess net energy savings, see the dedicated topical case study (Voswinkel et al., 2018).

11 | MESSAGE 10: Good data is well-documented data

> "The evaluation reports themselves are very detailed and are used as information sources for other programme evaluation reports (e.g. for the municipal Climate Protection Programme)."
>
> Quote from the case on the City EE programmes of Vienna (Austria)

> "In reality, if two persons carry out impact evaluation of the same policy measure, they get different results. Even if I make the same calculation in successive years without proper documentation of the calculation method and definitions, the calculation can be different. This highlights the needs for good logic and documentation."
>
> Quote from the case on Energy Efficiency Agreement for Industries (Finland)

In addition to the quotes above, this conclusion comes from the difficulties encountered about clarifying the data (scope, unit, etc.) when preparing the case studies. One of the objectives of the EPATEE case studies was indeed to document in a systematic way the data about energy savings and costs respectively. Figure 7 and Figure 8 (see section 1.4) show the variety of metrics used to report energy savings (about this issue, see also Voswinkel, 2018). The table below then provides a quick overview of the diversity in the scope of costs (when data about costs could be found).

Table 15. Feedbacks about the methods used to assess net or additional energy savings or impacts.

Types of cost data monitored	Cases where these cost data are mentioned
Amount of public aids (e.g., subsidies)	Environmental support scheme (AT), Primes Energie (BE), Individual heat metering (CR), Voluntary energy audits (FI), "Future investments" programme (FR), EE Fund (DE), Better Energy Homes (IE), Subsidy scheme for housing corporations (NL), Warm Front (UK), WAP (US)
Total investment costs of the energy efficiency actions supported	Primes Energie (BE), Energy renovation of public buildings (CR), Individual heat metering (CR), Voluntary agreements (FI), Voluntary energy audits (FI), EE Fund (DE), Better Energy Homes (IE), Warm Front (UK), WAP (US)
Annual savings on energy bills for the participants	Energy renovation pf public buildings (CR), Voluntary agreements (FI), Voluntary energy audits (FI)
Administration costs (for the public bodies in charge of the scheme)	EEO scheme (DK), Voluntary agreements (FI), Voluntary energy audits (FI), EE Networks initiative (DE), EE Fund (DE), Better Energy Homes (IE), White certificates (IT), Supplier obligations (UK), Warm Front (UK)
Total public costs of the scheme	Agreement for freight companies (FR), Warm Front (UK), Capacity markets (US), WAP (US)
Total costs for the obligated parties	EEO scheme (DK), White certificates (IT), Supplier obligations (UK)

Note: this table does not include types of cost data specific to one case.

The figure below illustrates the difficulty to find cost data in evaluation reports: apart from the amount of public aids (and stakeholders' costs when relevant), the other cost data could not be found in evaluation reports or other public reports for most of the cases. For more details, see section 9.2 of Volume II.

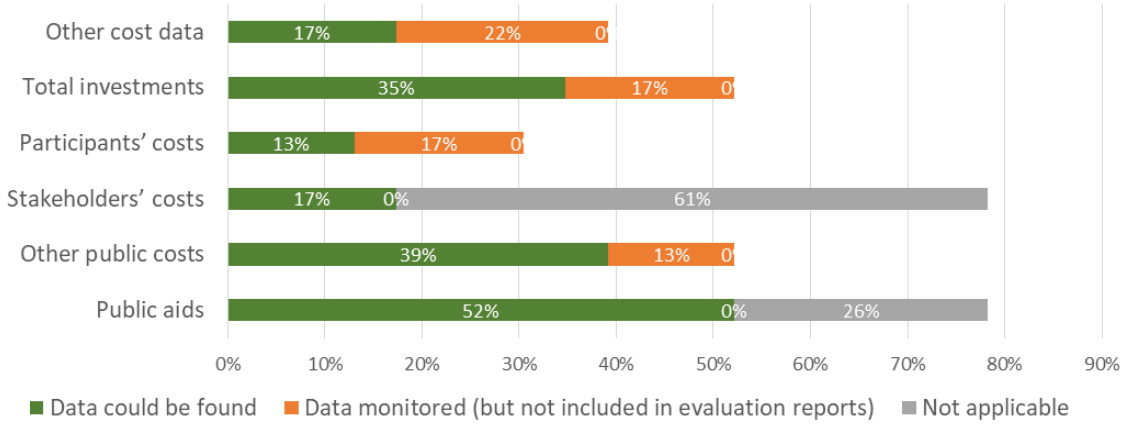


Figure 10. Availability of cost data in evaluation reports (or other public reports), per type of cost data (in % of the 23 case studies).

12 | MESSAGE 11: Communication about evaluation results can be as important as doing the evaluation

"The ex-post evaluation was very well accepted by the Ministry. Indeed the evidence brought by the evaluation changed their perception of the scheme. There were no more questions about the rationale or interest to implement this scheme. At the opposite, the questions were about how to make the scheme grow." Quote from the case on Better Energy Homes (Ireland)
"Sometimes, certain sentences need to be revised in their formulation or their content. In general though, discussions between the evaluators and the ministry are open and based on mutual understanding." Quote from the case on the Energy Efficiency Fund (Germany)
"It is always very motivating for us to know that the evaluation customer is genuinely interested in our results and that the results are used as basis for decisions regarding the policies that were evaluated. This also means that the stakeholders show a keen interest in the results. Our approach is when possible to engage the relevant stakeholders in the evaluation and to be open about the progress of the evaluation work so that the end result does not come as a shock. In my opinion, this approach benefits the overall decision-making process. (…) An additional advantage is that when we have had regular contact and discussion with the individual stakeholder groups, then all parties – us included – are better prepared for the media attention and can minimise the distortion of the evaluation results by the media. This is something which frequently happens." Quote from the case on the Energy Companies' Energy-Saving Efforts (Denmark)

The communication issue was not directly in the scope of the EPATEE case studies, but it comes out from several of the interviews. So it was not systematically covered in all case studies. This is one of

the subjects of the second phase of the EPATEE project, in the task about how to integrate evaluation into the policy cycle.

The interviews done for the case studies already made possible to identify the following key issues (for more details, see section 10 of the Volume II).

About the discussions of evaluation results and recommendations:

- The process to discuss them and how they will be used should be clarified from the start. This increases the motivation of the evaluators and stakeholders involved in the evaluation process, and avoid to generate frustrations.
- Intermediate discussions along the evaluation process are useful to avoid misinterpretations of the results and possible "shocks" (when unexpected results would be presented only at the end).

About the communication on the evaluation results and conclusions:

- Communicating evaluation results and conclusions to the main stakeholders is important for their involvement in the scheme (and in its monitoring and evaluation)
- The format of the evaluation results and conclusions should be adapted to the audience targeted. Likewise, if there is a need to raise the interest of a particular audience, this might influence the choice of the main indicators or messages to communicate.
- The main evaluation reports should be documented enough to enable a review of the results and keep the memory of the evaluation.
- Evaluation results should be presented in a way that evaluation users will know what caution to take when interpreting them. Evaluation conclusions should indeed clarify how the results should be understood.
- Communicating evaluation results to the general public can be a way to raise awareness and participation to the scheme.
- The choice of indicators reported might affect the key message to policy makers.
- Results and conclusions communicated to medias/press should be prepared carefully (to avoid conclusions to be misinterpreted or messages to be distorted)

13 | Beyond the evaluation of energy savings

The focus of the EPATEE project is primarily on the evaluation of energy savings. The feedback from the stakeholders' surveys clearly showed their interest also in the evaluation of other indicators, impacts or aspects. Therefore the case studies were also used to review what types of other indicators or impacts were evaluated together with the energy savings: types of indicators to assess cost-effectiveness and cost-efficiency (section 13.1), impacts other than energy savings (section 13.2) and other aspects (process evaluation, customer journey, participants' satisfaction, market transformation, etc. ; section 13.3) that have been evaluated for the policy measures analysed in the EPATEE case studies.

It should be noted that the EPATEE case studies did not intend to analyse the methods used to assess these other indicators or impacts, but to review which ones were evaluated.

13.1 Cost-effectiveness and cost-efficiency

The case studies for which information about cost-effectiveness of cost-efficiency could be found show a diversity in the related indicators (see table below). Moreover, the documentation of these indicators does not always make possible to know the scope and metrics of the benefits and costs taken into account (e.g., if energy savings were gross, net or additional).

More generally, even when the type of indicator seems very similar, results should not be compared directly, as there are most often differences in the scope or metrics of the benefits and costs taken into account, not to mention the differences in the methods used to assess these benefits and costs.

Table 16. Types of indicators used to assess cost-effectiveness or cost-efficiency.

Types of indicators used to assess cost-effectiveness or cost-efficiency	Cases where these indicators are mentioned
Public costs divided by lifetime-cumulated energy or CO_2 savings	Environmental support scheme (AT), Primes Energie (BE), Agreements for freight companies (FR), EE Fund (DE), Subsidy scheme for housing corporations (NL), Nordsyn
Obligated parties' costs divided by the reported energy savings	EEO scheme (DK)
Total costs of the scheme (including stakeholders and/or participants' costs) divided by lifetime-cumulated energy or CO_2 savings	Agreement for freight companies (FR), Better Energy Homes (IE)
Net Present Value of the scheme, considering different sets of costs and benefits according to different points of view	Better Energy Homes (IE), Supplier Obligations (UK)
Public expense for an additional electricity kWh saved	White certificates (IT) (until 2011)
Leverage effect	Environmental support scheme (AT), "Future investments" programme, EE Fund (DE)
Value for money	Warm Front (UK)
Net present value of additional energy efficiency actions (over their lifetime), taking into account related energy savings and emission reductions, and costs	EEO scheme (DK)
Various indicators at project/action level: Net present value of investment and energy savings at project level; Direct payback time; Action or project costs divided by its benefits; Savings to Investment Ratio (SIR) (lifetime energy bill savings divided by the costs)	Energy renovation of public buildings (CR), Individual heat metering (CR), Voluntary agreements (FI), Voluntary energy audits (FI), "Future investments" programme (FR), WAP (US)

Several interviews also pointed that cost-effectiveness indicators do not provide a complete view about the impacts of the scheme and how it works. They thus strongly recommend to consider other indicators, impacts or aspects (as illustrated in the following two sections).

13.2 Impacts other than energy savings

The figure below shows the type of impacts for which information could be found when preparing the case studies.

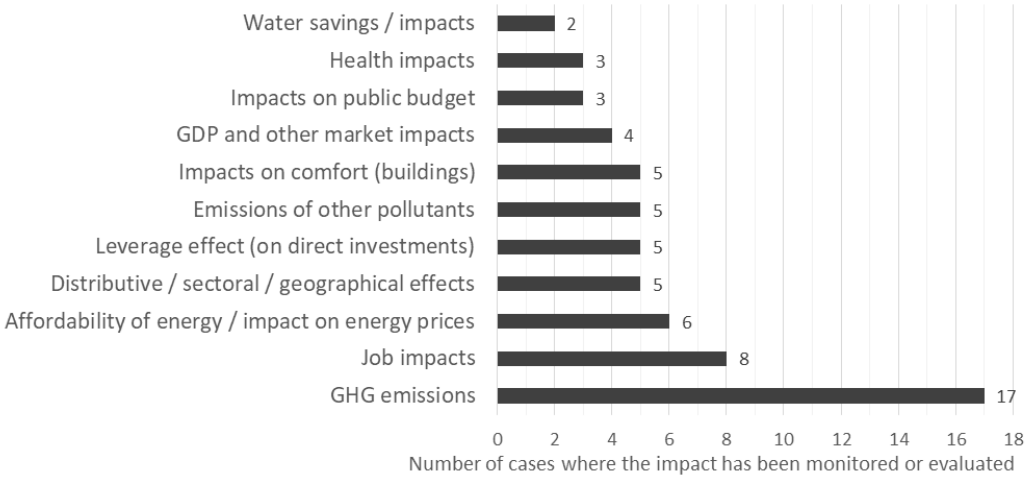


Figure 11. Frequency of other impacts in the 23 case studies, per type of impact.

Moreover, several interviewees pointed the importance to consider impacts other than energy savings (see quotes in the section 8.2 of the Volume II).

13.3 Other aspects than impacts

> "Evaluations are often focused on a limited set of indicators that might not be the most relevant. This is usually because funding for evaluation is often limited. Therefore, evaluation is ordered with a restricted approach and the main objective to assess whether the policy worked or not. However, this is only one part of what evaluation should be. Most of the time, it would be more important to understand WHY the policy worked or did not work"
>
> Quote from the case on Warm Front (UK)

The quote above highlights that evaluating the impacts of scheme provides only a partial understanding of its success. The figure below shows the number of case studies with information about three main other types of evaluation:

- process evaluation: when analyses were made with the objective of understanding how the scheme works and/or why results were achieved or not);
- customer journey / participants satisfaction: when analyses were made about the participants' point of view
- Market transformation: when analyses were made about the impacts of the scheme on markets related to energy efficiency products or services (e.g., availability of energy efficient products, trends in prices)

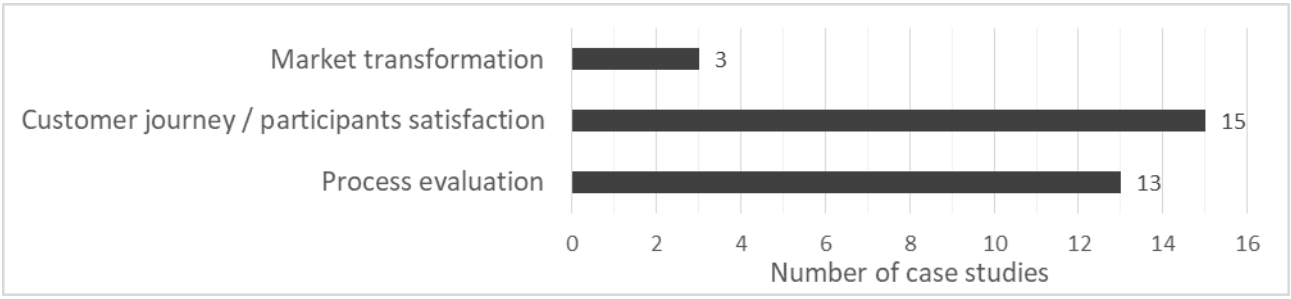


Figure 12. number of case studies with information about three main other types of evaluation

These results should be taken with caution, as looking at these aspects was not the primary focus of the EPATEE case studies, that reviewed mostly impact evaluations. Therefore, there could be other evaluations or studies on these cases that were not found when preparing the case studies.

Still, the figure shows a much lower frequency of analyses related to market transformation. Which is consistent with the fact that spill-over effects were evaluated in none of the case studies (see Figure 6 in section 1.4).

For more details, and especially for more specific information about the other aspects evaluated encountered in the case studies, see section 8.3 of the Volume II.

14 | Key points of debates

The interviews done for the case studies sometimes mentioned debates around evaluation issues, or provided different views (from different interviews) on the same issue.

This part provides a short summary about the debates on the three issues the most frequently discussed:

- independency of evaluators (or independent evaluation);
- relevance to assess net energy savings;
- choice between evaluating energy savings based on estimates (e.g., engineering calculations) or on measured or metered data (e.g., billing analysis).

About **independency**, the discussions were mostly about two aspects:

- what does independency mean: does it mean necessarily external evaluations? can external evaluators commissioned by the body in charge of the scheme be considered independent? can evaluators being from the same body but a different service be considered independent?
- pros and cons of internal/external evaluations: data access, confidence in the results and conclusions, legitimacy and credibility, critical analysis providing new insight

"Independence of evaluators is important but on the other hand evaluators have to know the scheme very well in order to understand the reasons for certain design choices policy makers made. This makes it sometimes difficult for authorities to find the right evaluators." Quote from the case on UFI - Aid for environmental protection measures (Austria)
"The issue of evaluators' independency should be considered in a pragmatic way. In the case of the Primes Energie scheme, the evaluation is done by a different service than the implementing service, but both services belong to the same department (Energy Department). On the one hand, one may say that this is not enough to ensure an independent evaluation. But on the other hand, this made that it was much easier for the evaluator to access the data needed for the evaluation, as programme managers may be reluctant to communicate data to persons outside their administration. They may have concerns about what the data will be used for, and about the way the evaluation conclusions will be drawn and communicated. Indeed, the adoption and ownership of the evaluation results require trust between programme managers and evaluators. So our experience is that the evaluation has more chances to be used, and recommendations to be implemented, when it is an internal evaluation."" Quote from the case on Primes Energie (Belgium – Wallonia)

For more details, see section 3.2 of the Volume II.

About **evaluating net energy savings**, the main discussion point is about whether it is possible to get reliable results. Indeed, ideal conditions are rarely met to use experimental or quasi-experimental methods (comparison between participants and a control or non-participants group). Therefore, when evaluating net energy savings, the most common approach in practice is to assess free-rider effects with surveys. This raised debates about the reliability of answers from the surveys.

"Finding a relevant control group to use quasi-experimental approaches (statistical comparisons of a participants and a control group) is often not possible and can be expensive. So, in many cases, the only option left is to use surveys. Therefore, evaluators at times have to rely on surveys where the respondents are self-reporting. Typically, the evaluation takes place some time after the activity and the end-users may not remember exactly what took place, why they chose to participate or not, what the cost was, and more importantly would they have acted differently if they had not participated. And finding the relevant person to talk to can be difficult, especially when it comes to non-residential consumers." Quote from the case on the Danish EEO scheme
"Counterfactuals are always difficult to define as there are no areas in the UK that have not been treated under the Supplier Obligation." Quote from the case on the Supplier Obligations (UK)
"More generally, it is very difficult to distinguish the effects of a measure from all the other factors that affect acting decisions. Particularly for measures that are already implemented for a long time. This is the case for the Primes Energie scheme that started in 2004. How to know what would have happened now, if the Primes Energie scheme had never existed? Meanwhile, the experience acquired with monitoring and evaluating the scheme over many years makes that we have some qualitative understanding about how it may affect acting decisions." Quote from the case on Primes Energie (BE)

For more details about evaluating net energy savings, see section 7 of the Volume II, and the dedicated topical case study (Voswinkel et al., 2018).

About **estimates vs. measured data**, the debates come from the large differences sometimes found between results from engineering methods and results based on energy bills (or other metered data). As shown in part 8 |, each type of method has its pros and cons, and each includes different sources of uncertainties. Therefore, there is no method that would be more reliable whatever the situation. It depends on several conditions: data availability, skills of the evaluators, etc.

About engineering methods, the main sources of uncertainties are often related to the assumptions used in the calculations, particularly about behaviours (see prebound and rebound effects) and

possible performance gaps. When using a model, a key point is whether the model could be calibrated on cases with metered or measured energy consumption.

> "There had been a kind of fairy story about the use of computer modelling in the 1980's and 1990's. With the development of computer capacities, there was an increasing use of modelling. However the ability to validate engineering data from modelling does not cope with measured ("real") data. Research of the last decade showed that it is essential to get "real" data to analyse complex systems such as energy use in buildings. Energy consumption cannot be explained by the description of the technical systems alone. There are interactions between technical systems, and above all between technical systems and occupants. This complexity is difficult to model."
>
> Warm Front (UK)

About billing analysis, the main sources of uncertainties are often due to sampling and matching methods. Depending on the cases, normalization for external factors might be needed, which can create other sources of uncertainties. Moreover, getting representative results do not only mean to have large enough samples, but also long enough time series. This might not fit the decision making timelines.

> "Two constraints make that ideal conditions for a perfectly robust evaluation are very rarely met: money, and often even more importantly time. If we take the example of evaluating impacts on heating consumption. This requires monitoring consumption over at least one heating season before and one heating season after the intervention. This already means a 2-year experiment. Then a third year is probably needed for the analysis and addressing issues in data collection, etc. But the evaluation customer normal wants results within one year, or less."
>
> Quote from the case on Primes Energie (BE)

The case studies also sometimes included some lessons learnt more specific to each case, or that did not correspond to the issues covered in this report. For more details, see section 11 of the Volume II.

15 | 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

EPATEE, 2017. Executive summary of the 1st European Peer-Learning Workshop. Available at: https://epatee.eu/events/1st-european-peer-learning-workshop-evaluation-practices-energy-efficiency-policies

Europe Economics, 2016. Evaluation of Fiscal Measures in the National Policies and Methodologies to Implement Article 7 of the Energy Efficiency Directive. Final report to the European Commission. Available online at https://ec.europa.eu/energy/sites/ener/files/documents/final_report_on_fiscal_measures_used_under_article_7_eed_0.pdf

Frederiks, Elisha R., Stenner, Karen, Hobman, Elizabeth V., and Mark Fischle, 2016. Evaluating energy behavior change programs using randomized controlled trials: Best practice guidelines for policymakers. *Energy Research & Social Science*, 22, 147-164.

Harmelink, Mirjam, Nilsson, Lars, and Robert Harmsen, 2008. Theory-based policy evaluation of 20 energy efficiency instruments. *Energy Efficiency*, 1 (2), 131–148.

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.

Labanca, Nicola, and Paolo Bertoldi, 2016. Energy Savings Calculation Methods under Article 7 of the Energy Efficiency Directive. Report of the Joint Research Centre for the European Commission. Available at: https://e3p.jrc.ec.europa.eu/publications/energy-savings-calculation-methods-under-article-7-energy-efficiency-directive

Maric, Lovorko, Thenius, Gregor, Gynther, Lea, and Catherine Guermont, 2018. Links between M&V tools (data collection) and evaluation (complementary analysis). Topical case study of the EPATEE project, funded by the European Union's Horizon 2020 programme.

ODYSSEE-MURE, 2015. Synthesis: Energy Efficiency Trends and Policies in the EU – An Analysis Based on the ODYSSEE and MURE Databases. September 2015. http://www.odyssee-mure.eu/publications/br/energy-efficiency-trends-policies-in-europe.html

Voswinkel, Fabian, 2018. Accounting of Energy Savings in Policy Evaluation – How to Get at Least 8 Different (Correct!) Results from the Same Data. Proceedings of IEPPEC 2018. Available at: http://www.ieppec.org/wp-content/uploads/2018/05/Voswinkel_paper_vienna.pdf

Voswinkel, Fabian, Broc, Jean-Sébastien, Breitschopf, Barbara and Barbara Schlomann, 2018. Evaluating net energy savings. Topical case study of the EPATEE project, funded by the European Union's Horizon 2020 programme.

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

Figures

Tables

Annex I: Terminology and typologies used in the EPATEE case studies

Terminology

1. Distinction between "measure" and "action"

In the EPATEE case studies, we chose to make a clear distinction between the terms "measure" and "action", in order to avoid the possible confusions when "measure" is used for multiple meanings:

- "**measure**" is used when speaking of policy measures (e.g., financial incentive schemes, EEO schemes, voluntary agreements, information campaign) ;

- "**action**" is used when speaking of the actions implemented at the end-user sites, and that can be either technical (e.g., replacement of a boiler, installation of insulation, replacement of lighting), organisational (e.g., implementation of an energy management system) or behavioural (e.g., eco-driving, switching off equipment to avoid stand-by consumption, efficient behaviours about ventilation).

2. Use of "ex-ante" and "ex-post"

The EPATEE case studies are focused on providing details and experience feedback about **ex-post evaluations** (= evaluations that were performed after the policy measure was implemented). We do acknowledge that an ex-post evaluation may make us of various data sources, and not only of data collected along or after the policy implementation. An ex-post evaluation may indeed use as well data that were estimated before the policy was implemented.

The minimum condition used in the EPATEE case studies to qualify an evaluation as "ex-post" is that the evaluation should take into account data monitored along the policy implementation and/or collected specifically for the evaluation about the number of actions actually implemented (and/or their effects as observed after their implementation).

At the opposite, an **ex-ante evaluation** is an evaluation that was done before the policy measure was implemented (or revised), and that by essence could not include data about the actual number of actions implemented.

We chose to **limit the use of "ex-ante" and "ex-post" to qualify evaluations, and NOT to qualify data**. Speaking of "ex-ante data" or "ex-post data" may sometimes be confusing. A distinction between "**data specific to the actions implemented**" and "**data taken from other studies or references**" can be more explicit.

Likewise, it may be confusing to speak of "ex-ante energy savings" and "ex-post energy savings". For example, how to name energy savings that are evaluated based on the actual number of actions implemented (as monitored along the policy) multiplied by standardised energy savings ratios defined as deemed savings? In this case, the number of actions would be ex-post, but the energy savings ratios would be ex-ante…

Therefore we chose instead to speak of "**expected energy savings**" when they were clearly assessed before the policy was implemented (e.g., for an impact assessment), and of "**reported energy**

savings" when the energy savings are the results reported about the policy implementation, and/or "**achieved energy savings**" when the energy savings were evaluated mostly based on data collected after the actions were implemented.

3. Other terms used in the EPATEE case studies

- "**Means**" = public budget committed, investments made by participants, costs for obligated parties and other financial means used to get the energy efficiency actions implemented.
- "**Outputs**" = number of participants, number of energy efficiency actions implemented (possibly detailed per type of action when relevant).
- "**Gross data**" = data taking into account all costs, investments, participants or actions monitored for the policy measure, and not taking into account possible free-rider or other effects related to the causality or attribution between the policy measure and the actions implemented.
- "**Net data**" = data estimated by taking into account a type of causality analysis between the policy measure and the actions implemented and/or by applying attribution rules (for example taking into account corrections for double counting between several policy measures, adjustments for free-rider effects, etc.). See below the typology and terminology used for correction and adjustment factors.

4. Frequent acronyms

The following acronyms are frequently used in the case studies:

- **Article 7 notification**: notification made by the Member States to the European Commission about the energy savings target required by the EED article 7, about the strategy to achieve this target
- **EED**: Energy Efficiency Directive (EU Directive 2012/27/EU)
- **EED art.7**: article 7 of the Energy Efficiency Directive (setting an energy savings target to be achieved either by the implementation of an Energy Efficiency Obligation scheme or by alternative policy measures)
- **EEO**: Energy Efficiency Obligation (as defined in the EED)
- **ESD**: Energy Services Directive (EU Directive 2006/32/EC)
- **EU**: European Union
- **NEEAP**: National Energy Efficiency Action Plan (plans to be reported by the EU Member States to the European Commission every three years, according to the ESD and then to the EED)

Typology used for policy instruments

We chose to use the same typology as defined in the MURE database (http://www.measures-odyssee-mure.eu/), to ensure consistency with already existing databased about energy efficiency policies in Europe:

- **legislative/normative** (mandatory standards, mandatory DSM, regulations on buildings, heating systems, vehicles, others)
- **legislative/information** (mandatory audits / energy managers / building certificates / labelling)

- **financial** (grants, subsidies, loans, others)
- **fiscal/tariffs** (eco-/energy-/CO_2-taxes; tax exemptions / reductions)
- **information/education/training** (energy billing, information campaigns, voluntary energy audits, regional/local information centres, voluntary labelling, others)
- **cooperative** (technology procurement, voluntary agreements, voluntary DSM measures, others)
- **Market-based** (energy efficiency obligations, energy efficiency auctions/tender systems, emission trading systems, JI/CDM)
- **Infrastructure** (only relevant for transport, e.g. modal shift, urban traffic management))

Typology used for the sectors

Likewise, we chose to use the same typology as defined in the MURE database for the end-use sectors:

- **Residential**
- **Tertiary** (+ possible to distinguish "public sector" or "private sector" when relevant)
- **Industry**
- **Transport**
- **Transversal** (used for measures that cover all end-use sectors listed above ; for measures that cover several but not all of these sectors, then the different sectors covered are mentioned; for example: "Household + Tertiary")

Other types of sectors were added to this typology as the case studies could have a broader scope:

- **Agriculture**
- **Energy sector** (energy generation, transmission or distribution) **Energy distribution** (+ possible to distinguish "electricity distribution", "gas distribution" or "district heating" when relevant)

"energy generation" includes policies about energy generation plants that are not "energy generation for self-consumption". For example, policies about CHP in industrial sites are included in "industry", and policies about domestic solar water heaters are included in "residential".

"energy transmission" includes policies about energy efficiency in electricity and gas transmission networks.

"energy distribution" includes policies about energy efficiency in district heating, electricity and gas distribution networks.

Typology used for the calculation methods

Two levels of typology are used. Level 1 is used to reflect the typology defined in the Annex V of the EED. Level 2 is the typology used in the MURE database.

Level 1 (= general categories defined in EED Annex V):

Categories	Explanations
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*"
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)

Level 2 (= types of methods used in the MURE database):

Bottom-up methods		*Link with level 1 categories*
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 categories*
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	other

Mix methods		*Link with level 1 categories*
	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	
Top-down methods		*Link with level 1 categories*
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 categories*
Method 10	Complex combinations of top-down and bottom-up methodologies in the form of integrated top-down and bottom-up methods	other

Typology used to describe the baselines

This typology was slightly adapted from the one defined in the MURE database:

- "**actual before**" energy consumption = metered energy consumption for the site, equipment, etc. where the energy efficiency action was implemented ;
- "**stock average**" = statistics available about the stock of buildings, equipment, etc. that make possible to estimate an average energy consumption of buildings, equipment, etc. in the stock ;
- "**before**" energy consumption = when the baseline can be either "actual before" energy consumption or "stock average"
- "**market average**" = statistics available about the buildings, equipment, etc. sold on the market and that make possible to estimate an average energy consumption of buildings, equipment, etc. sold on the market ;
- "**control group**" = energy consumption of households, companies, etc. that form a control group for a statistical analysis
- "**trend**" = e.g. consumer behaviour, autonomous technical progress, "business as usual" scenario
- "**minimum or performance standards**" = only savings exceeding standards are accounted
- "**other**" = in case the baseline used does not correspond to any of the types listed above (then provide a definition of this type of baseline in the case study)

Note: several types of baselines may be combined depending on the evaluation method used.

Typology used for adjustments, correction and other factors

Whenever possible and needed, the terms and definitions used in the evaluation reports are mentioned in the case studies. In addition, the following typology was used to provide a common

basis for the analysis of the case studies, and for consistency with the terminology used in the EPATEE Knowledge Base:

- **Free-rider effect**: a free-rider is an end-user who beneficiated from the support (energy advice, financial incentives, etc.) provided by the measure, but who would have implemented the energy efficiency action in the absence of the measure[1].
- **Spill-over effect**: energy savings caused by the measure and other than those resulting from participants implementing energy efficiency actions promoted by the measure[2].
- **Direct rebound effect**: "Improved energy efficiency for a particular energy service will decrease the effective price of that service and should therefore lead to an increase in consumption of that service. This will tend to offset the reduction in energy consumption provided by the efficiency improvement"[3].
- **Indirect rebound effect**: "The lower effective price of the energy service may lead to changes in the demand for other goods, services and factors of production that also require energy for their provision. For example, the cost savings obtained from a more efficient central heating system may be put towards an overseas holiday"[3].
- **Prebound effect**: cases where, before implementing an energy efficiency action, end-users tend to consume less energy than estimated by engineering models[4].
- **Performance gaps**: cases where the observed energy performance of the energy efficiency action installed is lower than the expected energy performance, for example due to differences in operating conditions or due to quality issues like defects when installing the action.
- **Multiplier effects**: imitators inspired by the facilitating measure but not benefiting from it.
- **Double-counting / measure interaction**: measures aiming at the same target can interact; i.e., one end-use action can be influenced by two or more energy efficiency measures.
- **Non-compliance**: stipulations are not fulfilled (esp. in case of regulatory instruments).
- **Others** (please specify)

[1] Several types of free-rider effects may be distinguished when considering the time when the end-user would have implemented the action in the absence of the measure ("deferred free-ridership") and/or the level of energy performance of the action that the end-user would have implemented in the absence of the measure ("partial free-ridership"). For more details, see SRCI, NOVEM, Electricity Association, MOTIVA, et al., 2001. A European Ex-Post Evaluation Guidebook for DSM and EE Service Programmes. SAVE Project No. XVII/4.1031/P/99-028, April 2001. (p.66)
http://www.evaluate-energy-savings.eu/emeees/downloads/Ex-post_Eval_Guidebook_DSM.pdf

[2] Several types of spill-over effects may be distinguished, in particular participants' spill-over (when participants implement other energy efficiency actions than the ones promoted by the measure) and non-participants' spill-over (when the measure has influenced end-users to implement energy efficiency actions, but who did not receive support from the measure). For more details, see SRCI et al., 2001 (p.69) (see complete reference in previous footnote)

[3] Sorrell, S., Dimitropoulos, J., 2008. The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics*, 65(3), 636-649.

[4] See for example: Sunikka-Blank M, Galvin R., 2012. Introducing the prebound effect: the gap between performance and actual energy consumption. *Building Research & Information*, 40(3), 260-273.