A study of the comparability of energy audit program evaluations

Elias Andersson, Oskar Arfwidsson, Victor Bergstrand and Patrik Thollander

Journal Article

Tweet

LINKÖPINGS UNIVERSITET

A study of the comparability of energy audit program evaluations

Elias Andersson*, Oskar Arfwidsson, Victor Bergstrand, Patrik Thollander

Department of Management and Engineering, Division of Energy Systems, Linköping University, SE-581 83 Linköping, Sweden

*Corresponding author. E-mail address: elias.andersson@liu.se (E. Andersson).

Amount of words: 7316

Abstract

There is a large untapped potential for improved energy efficiency in various sectors of the economy. Governmental industrial energy audit programs subsidizing the companies to conduct an energy audit are the most common policy in trying to overcome the energy efficiency gap. Evaluation studies have been carried out to gain knowledge about the success of a completed energy audit policy program. The evaluations were made in different ways and in addition focused on different performance indicators and used different ways of categorizing data. In this article, a literature review has been made of five evaluation studies from different energy audit programs, where the problems of the present incomparability between programs due to differences are discussed. The policy implication of this paper is that new energy audit policy programs must specify a harmonized way of categorizing data, both regarding energy efficiency measures and energy end-use. A proposition for a standard for how to evaluate energy audit policy programs is also described in the article. Conclusions from this study are that important elements, such as the free-rider effect and harmonized energy end-use data, should be defined and included in evaluation studies. A harmonized standard for evaluating energy audit programs is not least needed within the EU, where member states are obliged to launch audit programs for large enterprises, and preferably also for small and medium-sized enterprises. This paper serves as an important contribution for the development of such a standard in further research.

Keywords: Energy efficiency, Energy audit policy evaluations, Industrial energy audit policy programs

1. Introduction

There is a large untapped potential for improved energy efficiency in various sectors of the economy. Governmental industrial energy audit programs subsidizing the companies to carry out an energy audit are the most common policy in trying to overcome the energy efficiency gap (Bertoldi, 2001). Energy audits are useful tools for identifying energy efficiency opportunities and energy saving potentials (Basurko et al., 2013; Petek et al., 2016). Energy audit policy programs therefore help companies to become more energy efficient, which in turn mitigates CO_2 emissions (Bertoldi, 2001; Fleiter et al., 2012b). It is important to compare the results of these energy audit policy programs to allow policy makers to design energy audit programs as cost effective as possible (Price and Lu, 2011). However, there seem to be many different ways in which these programs are conducted and how they have been evaluated (Fleiter et al., 2012a; Price and Lu, 2011), which is why an international comparison can be difficult to accomplish. Issues connected to the evaluation of energy efficiency programs, namely technical, policy, and infrastructural issues, within the US context were analyzed by Vine et al. (2012), who highlights that differences in e.g. definition of energy impact or free-rider effect complicate the comparability between programs. Farla and Blok (2002) also note difficulties in comparing energy end-use data from Long-Term Agreements (LTAs) regarding energy intensity development with national statistical data in the Netherlands due to large differences.

In order to adequately assess the successfulness and evaluate the outcome of an energy audit program with the aim to compose future appropriate national energy audit programs, it is important that the evaluations are conducted in a harmonized way, especially regarding performance indicators and elements such as cost effectiveness, energy savings and free-rider effect. Thomas et al. (2012) raise the need for harmonized methods for evaluation of energy savings from energy efficiency programs and propose methods to be used depending on the quality of data. However, in order to facilitate evaluation of energy savings and cost effectiveness of energy efficiency programs, particularly programs focused on energy audits, there is a need for both consistency and transparency in how the evaluations are made and how the energy efficiency measures (EEMs) are categorized.

The aim of this study was therefore to evaluate the comparability of evaluation studies of energy audit policy programs. Energy audit policy program evaluations were studied to investigate comparability issues such as inconsistency in evaluation and how the EEMs are categorized. It is important to keep in mind that the programs include a number of elements which can differ from each other (Väisänen et al., 2003), including life span, amount of subsidy, type of industry (e.g. small and medium-sized enterprises[1] (SMEs) or energy intensive), country where it is conducted, and number of participants. These differences might affect the comparability of the evaluation studies in different ways. The structure of the paper is as follows: an overview of program evaluation theory is presented first. Second, the method used for choosing which energy audit programs to study and how to analyze the comparability between them is explained, followed by a brief presentation of the chosen energy audit policy programs. Third, a section with the results of how the programs have been evaluated followed by a discussion of the findings. The paper ends with conclusions and policy implications.

2. Public Policy Evaluation

One role of evaluation in the context of public policies is to assess the impact of a policy program. The basic procedure, as Weiss (1998) explains, is to compare participants before the program is enacted and when the program is finished. This analytical process is especially relevant for public policy makers, as evaluations of old programs are needed to plan and design more effective future programs. Vedung (1997) defines evaluation as "*careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of governmental interventions, which is intended to play a role in future, practical situations*". Program evaluation is thus an important part of the lifecycle of public policy programs in order to develop programs with a more desirable outcome. The overarching point of interest as regards evaluation is to improve the performance of public policy and public interventions. An important part of evaluations

---

[1] The definition of SMEs is usually defined by number of employees and varies depending on in which country the program has been conducted.

to ameliorate, as investigated by Vine (2008), is that the presentation of evaluation results is, among other things, specific and relevant to the needs of the program.

Apart from evaluating a policy program, the evaluation itself is as auditable as any other management process. This type of evaluation might be conducted by the evaluator him- or herself, where aspects such as methodological quality, readability, and care to stick to facts should be included in the investigation (Vedung, 1997). Furthermore, if a number of evaluations of similar kinds of programs have been made, it is possible to conduct a so-called *meta-analysis* or *meta-evaluation*, where data from multiple programs are put together and generalized (Weiss, 1998). Weiss (1998) mentions a few important purposes of doing meta-analysis, one of them being to investigate under which conditions the most desirable outcomes from a program are achieved. If successful, the meta-analysts might use the data from the different studies to map what actions and settings provide satisfactory outcomes more effectively than if only one evaluation was studied. For policy program designers, who cannot possibly acquire all available outcomes and results from different policy program evaluations, achieving outcomes generalized from multiple evaluations through a meta-evaluation can be of great value, e.g. under what conditions are more preferable results achieved? A meta-evaluation might offer a clearer explanation of effects than a single evaluation study might (Weiss, 1998). Weiss (1998) argues that the main advantage of a meta-evaluation is the improvement in estimation of program outcomes.

Weiss (1998) mentions a few requirements that need to be fulfilled in order to conduct a meta-analysis: the programs should have common goals, similar measures for the outcome, and include similar activities. Therefore, in order to realize a meta-evaluation of energy efficiency programs, the different evaluation studies must contribute enough and adequate data as well as harmonized categorization of data. Stenqvist and Nilsson (2012) believe that the most ideal procedure is to develop a plan for evaluation in the early stages of designing a policy instrument. This would most likely reduce the number of assumptions needed in the evaluation and allow an integration of relevant program data.

3. Method

Five evaluation papers of energy audit programs from Australia, Germany, Sweden and the United States have been studied in depth. These papers were chosen as they represent well-known energy audit program evaluations from different countries and have been cited multiple times. The methodology follows a multiple-case holistic design, as explained by Yin (2014), where the studied papers are used for analytic generalization of the proposed hypothesis – that a harmonization of elements is needed to achieve comparability between evaluations. The findings are then assessed in the context of policy implications, where an analysis is made of how to facilitate for public policy makers to use the comparison of energy audit program evaluations to design future cost effective programs.

Still, other evaluations could potentially have been included in this study, but these five evaluations were deemed by the authors to fulfill the purpose of the paper, displaying what differences or similarities exist in various evaluations. However, additional studies of other energy audit program evaluations might not only enhance the results from this study by further assess the comparability and the need for a harmonized standard for categorizing energy end-use data, but also develop further guidelines for how to evaluate energy audit policy programs.

The elements that were investigated in the chosen evaluations were whether financing of energy audit, public and firm cost effectiveness, discount rates, free-rider effect, control group, energy savings, and CO_2 reduction were included. If so, the way the results were presented in the program evaluations in terms of performance indicators, payback time, and categorization of implemented measures were compared. In addition to this, the data collection method used in the evaluation studies is presented.

4. Energy audit policy programs

Anderson and Newell (2004) evaluated an energy audit program in the US, the Industrial Assessment Centers (IAC) program. The program included more than 10,000 energy audits from between 1981 and 2000. The participating firms had less than 500 employees and an annual energy bill of between US$ 100,000 and US$ 2,000,000.

In Australia, a program called Commonwealth Government's Enterprise Energy Audit Program (EEAP) operated between 1991 and 1997 (Harris et al., 2000). Around 1,200 firms participated, most of them non-SMEs with an average of 297 employees.

In 2008, an energy audit program for SMEs was introduced in Germany and the first two years of the program were evaluated by Fleiter et al. (2012a). The program offered grants for energy audits that covered most of the cost. 10,400 audits were approved during the evaluated period; of these, 80% were initial audits and 20% comprehensive audits. About 50% of the participating firms had fewer than 20 employees while the mean participating firm had 38.

The industrial energy efficiency program Project Highland was conducted in Sweden between 2003 and 2008, focusing exclusively on SMEs (Thollander et al., 2007). The firms were each given an energy audit, although no complete audit was made due to the risk of unfair competition. Therefore, the auditors were restricted to conducting the energy audits during two days; the first day to be spent on a field visit and the second on compiling the data and writing their report.

While Project Highland was performed in a certain region in Sweden, a follow-up national program directed at Swedish industrial SMEs was adopted in 2010 to give the companies energy audit checks (Thollander and Dotzauer, 2010). SMEs using more than 500 MWh annually were included and the program was set to continue for five years until 2014 with the aim of conducting energy audits for 900 companies. Companies could apply for financial support for 50 percent of the cost of an energy audit, up to a maximum of €3,000[2] (Lublin and Lock, 2013).

5. Energy audit policy program evaluation

When the energy audit policy programs were evaluated, the results may have been affected by the level of detail of the energy audits made in the different programs or even at an individual firm. ISO 50002 states that "*Organizations may adjust the level of detail of the energy audit between type 1 and type 3 to suit the*

---

[2] In order to compare figures, 1 euro was assumed to be equivalent to 10 SEK.

needs of the organization." (International Organization of Standardization, 2014), where the different types basically imply different levels of details in the audit made. The level of detail in the audits was seldom explained in the published policy evaluation studies, and this should therefore be kept in mind when analyzing their results.

The IAC study was conducted by analyzing records stored in the IAC database. In the assessment of the program, the EEMs were classified into eight different types, for example *building and grounds*, *motor systems*, and *thermal systems* in order to capture the heterogeneity in different type of measures. The authors used estimated implementation cost and annual operating cost savings from the IAC database to calculate the payback time for each EEM. Free-riders were not taken into account. The average energy price was calculated for each EEM by dividing annual savings by the quantity of saved energy. The prices of different energy types were normalized to have a mean value of one (1) in order to make the prices comparable. In the evaluation study of the IAC, implicit discount rates were calculated assuming a project lifetime of ten years. Therefore, the rate was calculated for each measure implemented. Firm cost effectiveness was presented in saved US$ and total energy savings were presented in Btu/year. No indicators of program cost effectiveness or reduction of CO_2 were presented.

The method used to evaluate the EEAP was to interview 100 of the participating firms, where each interview was made by telephone and lasted around 30 minutes. Before the interview a number of questions had been prepared, which were denominated either as g*eneral information questions*, *decision rule questions* or *sources of risk questions*. The evaluation was made two years after the program ended. The suggested measures were divided into 14 categories, ranging from *lighting* and *air conditioning* to *boilers* and *steam production*. The most common recommendations were categorized as *lighting* and *general suggestions*. When calculating pay off times for the suggested measures, six different discount rates were used: 6, 8, 10, 12, 15, and 20%. To calculate the cost effectiveness for each participating firm, the evaluation study put together the average value of implementations made by the companies and calculated a mean total net present value for different discount rates for all the firms' implemented measures, and presented them in one table. However, no categorization of type of implementation was made. From the figures presented,

average savings for an implementation depending on investment lifetime and discount rate can be obtained in A\$. However, no information regarding energy savings in kWh for either the firms or the public was provided in the evaluation study, and no account was taken of possible free-rider effects. No CO_2 reduction was presented either.

In the evaluation of the German energy audit program, the data was collected from an online survey among the firms with a response rate of 12%. In addition to this, 107 audit reports and statistics from the KfW, the German bank that managed the project, was analyzed. The participating firms were not required to report achieved energy savings, and the energy savings resulting from the program were therefore calculated from the results of the survey. The number of measures implemented in each firm was not included in the audit reports but was cross-checked with the survey. The net energy savings were calculated with and without taking free-rider effect into account, which was stated through a question in the survey of which EEMs were definitely planned to be implemented, and the results for the entire program were extrapolated from the sample. Many EEMs had not yet been implemented due to the short time between the audit and the evaluation, therefore the evaluators assumed that 50% of the measures that would "possibly" be implemented were implemented and all measures that would "certainly" be implemented were implemented. Because of the uncertainty in the lifetime of EEMs and future energy prices, a lower and an upper boundary of the profitability were calculated. The net present value from the measures was calculated with a firm, public, and societal perspective with assumed discount rates of 15, 3, and 3 %, respectively. In the calculations, EEMs that were certainly planned before the audits were excluded to correct for the free-rider effect. This was stated through a question in the survey. The measures were divided into 13 different types. For each type of measure, different mean project lifetimes were used, ranging from three to thirty years. The authors presented firm cost effectiveness in €/MWh and €, program cost effectiveness in €/MWh, total energy savings in PJ/year, and CO_2 reduction in kt CO_2/year.

The energy audits for Project Highland were conducted by Energy agency of South East Sweden (ESS). A questionnaire with individual questions for each firm was prepared to evaluate the project. This questionnaire was sent by mail and in some cases complemented by interviews. The different measures

were divided into eight categories. One was called production processes and included all types of production-related processes, while the remainder were related to support processes such as *space heating*, *lighting*, etc. It is important to bear in mind that the program was directed at non-energy-intensive firms and SMEs. This might affect the total number of measures as well as the number of categories of suggested measures, in the sense that the firms' main energy usage might not be from the production processes. In this evaluation study, an energy saving potential for the whole program was presented in terms of energy saved per year for both electricity and other energy carriers (in MWh/year). Furthermore, the evaluation study presents saved energy per public euro invested in the program in kWh/€ To estimate the free-rider effect for Project Highland, Thollander et al. (2007) asked the six firms with the highest adoption rate and the five firms with the lowest adoption rate if they would have implemented the EEMs even without the energy audit. No indicators of firm cost effectiveness or CO_2 reduction were presented. No payback times for the firms were calculated and therefore no discount rate was used.

For the energy efficiency program, where Swedish industrial SMEs were given energy audit checks, an ex-ante evaluation was made for the program's cost effectiveness (Thollander and Dotzauer, 2010). To estimate the free-rider effect, Thollander and Dotzauer (2010) refer to the study by Thollander et al. (2007), where firms state that the measures would not have been implemented without the audit. They further address the free-rider effect that the companies would have made an energy audit even without the program by relying on a previous study by Väisänen (2003), who argues that this specific free-rider effect is very low for industrial energy audit programs. Based on the evaluation of Project Highland by Thollander et al. (2007), the public costs have been estimated for the program (in kWh/€), as well as a total estimation of potential energy savings (in MWh/year). Even though the program differs from Project Highland since the audit subsidy covers only 50% of the cost, an estimation could be made. The degree of a possible free-rider effect was taken into account through a comparison with Project Highland and studies of other European programs. However, as in Project Highland, no firm cost effectiveness or CO_2 reduction was presented.

None of the evaluation studies used a control group not affected by the energy audit program to assess the impact difference. Fleiter et al. (2012a) mention the use of a control group, but argue that they could not do

so due to too low a response rate from the companies in the program. How each of the evaluation studies presented its performance indicators and other elements is summarized in Table 1.

Table 1. Comparison between different energy audit program evaluation studies

Evaluation study; program; country	Time span [years]	No of parti-cipants	No of measure cate-gories	Life-time of EEMs [years]	Discount rate used [%]	Free-rider effect taken into account	Control group used	Data collection method	Performance indicators			
									Firm cost effective-ness	Program cost effectiveness	Total energy savings	CO_2 reduction
Anderson and Newell (2004); IAC; USA	20	9034	8	10	Implicit discount rate	No	No	Database	Yes [US$]	No	Yes [Btu/year]	No
Harris et al. (2000); EEAP; Australia	6	1200	14	1-10	6, 8, 10, 12, 15 and 20	No	No	Questionnaire over tele-phone, strati-fied random sampling	Yes [A$]	No	No	No
Fleiter et al. (2012a); Germany energy audit program; Germany	2	10400	13	3-30[a]	15 (firm), 3 (public), 3 (society)	Yes[b,c]	No	Online survey, audit reports and program statistics	Yes [€/MWh] and [€]	Yes[d] [€/MWh] and [€]	Yes [PJ/year]	Yes [kt CO_2/year]
Thollander et al. (2007); Project Highland; Sweden	5	340	8	-	-	Yes[b]	No	Questionnaire and interviews	Yes [kWh/€]	Yes [kWh/€]	Yes [GWh/year]	No
Thollander and Dotzauer (2010); Swedish energy audit program; Sweden	5	900	-	-	-	Yes[c]	No	Estimated cost effectiveness based on other programs calculations	No	Yes [€/kWh]	Yes [GWh/year]	No

a) Depending on type of measure.

b) Free-rider effect meaning that the companies would have implemented the EEMs even without the program.

c) Free-rider effect meaning that the companies would have made an energy audit even without the program.

d) Calculated cost for both public and society.

6. Discussion

The results of this study show that there are differences in how the evaluation studies of the energy audit programs were conducted and how the results are presented. Even if a single evaluation has been clear in methodology and presentation of results, the comparison between the evaluations shows that number of categories as well as type of categorization of measures differ between the studies. It is clear that there is lack of consistency in how the measures are categorized. Only four out of 24 categories are common to Fleiter et al. (2012a), Harris et al. (2000) and Thollander et al. (2007), as shown in Figure 1. The taxonomy of EEMs by Anderson and Newell (2004) differed most and did not have any category in common with the other evaluation studies. Regarding the study by Thollander and Dotzauer (2010), as they made an ex-ante evaluation of a program, the results investigate only the cost effectiveness based on a number of studies (Thollander et al. (2007) among others), hence do not present any categories of measures. As a consequence, the energy efficiency and cost saving potentials for one category in one program cannot be compared to the potentials of the same category in another program. Furthermore, none of the studies presents a definition of what a measure is and therefore the number of measures presented in the evaluations cannot be compared either. For example, one evaluator might categorize lighting control and replacement of fluorescent lamps as one single measure, while another evaluator categorizes them as two different measures. It might even be categorized by an evaluator as one behavioral measure and one lighting measure.

The performing of an energy audit affects the categorization of measures and availability of data. Such a categorization should be made prior to the energy audits being conducted to facilitate for policy evaluation studies to use common categories for measures. ISO 50002, which has entered into force, does not mention how the categorization of EEMs and energy end-use data should be made. In fact, there is a lack of common ways to categorize EEMs for SMEs, as also noted by Thollander et al. (2015). In light of the strategic attempt by the European Commission to try and harmonize data for the building sector (EEFIG, 2015), a similar work for industrial companies would be highly needed.

ISO 50002 addresses the possibilities for each auditor to adapt the level of detail of the audit between type 1 and type 3 to suit the needs of the company, where a more detailed level (type 3) generally requires more measurements. As this may be flexible for auditors and companies, it may at the same time affect how much data is available, where some audits are more comprehensive and provide more energy data than others. Only Thollander et al. (2007) mention the level of detail of the audits in the program, but none of the evaluation studies describes the level applied thoroughly. If reported, information regarding the level of the energy audits could increase the comparability of the studies.

Further, a comprehensive energy audit on a detailed level is to be preferred as it can bring additional benefits for individual industrial companies as well, inspiring them to work continuously with energy efficiency by incorporating it into the company's or organization's management. Performance indicators connected to energy end-use could then be developed as energy-related key performance indicators (KPIs) to be monitored. In an industrial company's work with energy management, formal management controls, such as measuring performance, are found by Schulze and Heidenreich (2016) to strengthen energy efficiency. Schulze et al. (2016) mention controlling and monitoring of energy use, including the use of relevant KPIs, as one of five key elements in an integrative work with energy management.

Table 1 shows that the performance indicators used in the studies differ. All studies have used cost effectiveness as an indicator of performance, either for firm, program or both. If a comparison of cost effectiveness between energy audit policy programs is to be made, the performance indicators should match. Fleiter et al. (2012a) presented total cost savings per firm as well as cost savings for each saved MWh, while Harris et al. (2000) and Anderson and Newell (2004) presented only the total cost savings per firm. This indicates that even if firm cost effectiveness has been calculated, it has been presented in different ways.

Furthermore, regardless if whether the performance indicators match or not, results show that the way in which they have been calculated differs. For example, Fleiter et al. (2012a) calculated firm cost effectiveness using investment costs, firms' share of the audit costs, and energy cost savings, while Harris et al. (2000) used the costs and benefits provided by the auditors from the investments implemented by the

firms, although not explicitly explaining how the calculations were made. Thollander et al. (2007) calculated the private capital spent by dividing the total energy saved by the companies' investment costs. In addition, different discount rates as well as different lifetimes for EEMs have been used in the programs, which affects the firm cost effectiveness. Thollander and Dotzauer (2010) did not calculate firm cost effectiveness, which means that this study cannot be compared with the other studies in respect of this performance indicator. Instead, they presented a program cost effectiveness, which was estimated based on the evaluation of Project Highland made by Thollander et al. (2007), who calculated the program cost effectiveness by dividing the total energy saved by the total amount of subsidy and administration costs. While keeping in mind that Thollander and Dotzauer (2010) made an ex-ante evaluation, the estimation differs from the calculation made by Fleiter et al. (2012a), who estimated the total public costs by adding together the program costs (monitoring and the processing of application, etc.) and the public share of the audit costs, also including the public discount rate and the lifetime of EEMs.

Table 1 also shows that the total energy savings of the programs have been presented in four of the evaluation studies. In three of these studies, program cost effectiveness has been calculated as well. The Swedish government emphasizes that an energy efficiency program should have a certain cost effectiveness in order to conduct it (Thollander and Dotzauer, 2010) and program cost effectiveness is therefore a vital performance indicator for policy makers. Furthermore, even though this study applies a bottom-up approach, it is worth mentioning that none of the evaluation studies corrected for structural effects.

As Weiss (1998) points out, to be able to conduct a meta-analysis of different programs they need to have a certain degree of similarity as regards goals, measures, and activities. Therefore, in order to achieve evaluation and comparison of multiple energy efficiency programs, with common goals and measures, a general standardization of program evaluations should be pursued. Vine et al. (2012) have stressed the importance of a national protocol in the US in the pursuit of higher quality standards for evaluations of energy efficiency programs. They emphasize that a national protocol should include common elements for e.g. definitions, methods, saving values, and formats of reporting results. In addition, Vine et al. (2012) mention that the use of consistent elements potentially allows energy savings from different evaluations to

be compared. This should be sought within the EU as well, where a standard and a protocol for evaluation of energy audit programs would greatly help increase comparability between program evaluations. Such a standard would benefit from also including guidance for how to plan for evaluation already in the design phase of an energy audit program, as mentioned by Stenqvist and Nilsson (2012).

Another important aspect regarding program cost effectiveness is the free-rider effect. Table 1 shows that three of the evaluation studies have taken this effect into account, but in different ways. Thollander et al. (2007) take measures that would have been implemented even without the energy audit into account, while Thollander and Dotzauer (2010) take the free-rider effect into account in the sense that the energy audit would have been made even without the programs. Fleiter et al. (2012a) address both these types of free-rider effect. In the other two studies, the free-rider effect was not mentioned at all. Fleiter et al. (2012a) claim the effect of EEMs that would have been implemented even without the program to be 20.9% and the amount of audits that would have been made without the program to be 9.1%. Thollander and Dotzauer (2010) argue that research has shown the free-rider effect to be very low. Thollander et al. (2007) claim the free-rider effect to be zero, however. The free-rider effect is important to estimate since it affects the cost effectiveness of the program, which in turn is an essential indicator for policy makers. The free-rider effect could have been included in those evaluations where interviews or surveys were used by simply adding a question. This paper reveals both a need for this to be included, but also for further research on how to properly estimate the free-rider effect.

Both methods for estimating the free-rider effect have their advantages and disadvantages. Making the assessment by asking the industrial companies if they would have made an energy audit even without the program (and the subsidy) tells us how many energy audits were made thanks to the specific program. However, estimating the free-rider effect by how many of the EEMs that still would have been implemented without an energy audit opens up the possibility of a more detailed assessment where an aggregated calculation of the cost-effectiveness of the program can be made down to each EEM. This means that a higher-quality estimation of the free-rider effect is attainable, but is also more demanding for both auditors and companies as well as entailing a risk of arbitrary assessments, e.g. a company assert they would

implement an EEM, while in reality they actually would not. With this said, which method would be most appropriate to determine the free-rider effect should be investigated in further research, but most importantly it needs to be comparable across different energy audit program evaluations.

In cost-benefit analysis, there are a number of methods for determining otherwise unmeasurable values, such as the free-rider effect, where all the methods are categorized as either *revealed preference* or *stated preference* (Mattsson, 2006; Kaval and Baskaran, 2013). The revealed preference method determines the value through observing human behavior, i.e. reveals desired information without directly asking for it (Mattsson, 2006). The stated preference method on the other hand is based on actually asking individuals the value of an element by means of interviews or questionnaires.

The most straightforward way of estimating the free-rider effect may be by including a question in a questionnaire or in an interview session (stated preference). However, estimating the free-rider effect through direct questions in interviews or questionnaires might risk being deceptive in that respondents claim that they, in all honesty, did not plan to conduct an energy audit if it were not for the policy program. This might be true at the time when the program was first initiated, but as factors change over time, e.g. energy prices rise, this first statement might be shifted towards another in the end; that an energy audit was actually going to be conducted, only later than at the time of the program's initiation.

Either way, it can be difficult to find a suitable control group to estimate the impact difference, as argued by Vedung (1997), who implies that this experimental design is more or less impossible to attain in the evaluation of a policy program. It could be argued that the programs could have used a control group consisting of eligible companies for economic support but that did not participate, but these already differ in their choice to not participate from the beginning, which proves a lack of interest in pursuing energy efficiency. Finding a physical copy of a group with exactly the same conditions as the targeted group, who are also unaffected by other political, economic, and social factors, will most likely be a fruitless search. This does not only apply to the free-rider effect, but to other elements in an evaluation as well.

Table 2 presents an overview of what should be included in an evaluation study depending on its level of detail. The suggestion is made from policy makers' point of view in order to facilitate for them to adequately

assess the successfulness of different energy audit policy programs and compare their cost effectiveness. Thus, this also implies that an evaluation study should at minimum include an estimation of the free-rider effect and a calculation of program cost effectiveness. However, standard definitions of the different elements are needed, especially for the free-rider effect, which this study shows is handled in different ways.
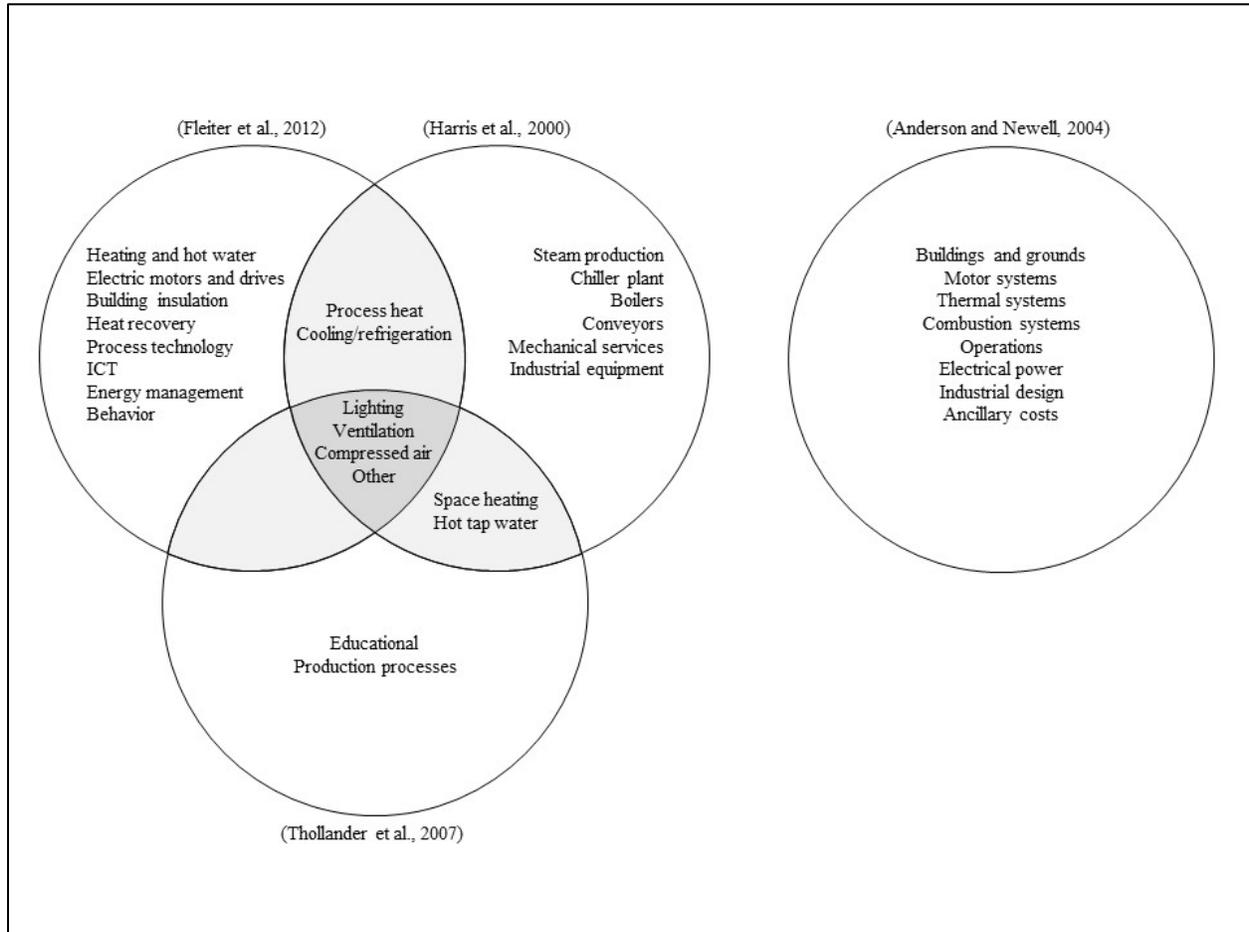

Figure 1: Diagram showing the different and common categories for the four studies by Fleiter et al. (2012), Harris et al. (2000), Thollander et al. (2007), and Anderson and Newell (2004). At the intersections the categories that the studies have in common are listed, e.g. lighting is a category in all of the studies while the category space heating is used only by Harris et al. (2000) and Thollander et al. (2007). None of the eight categories presented in the study by Anderson and Newell (2004) was common to any of the other studies and these are therefore shown as a distinct sphere. The study by Thollander and Dotzauer (2010) did not present any categorization of measures, why that study is not present in this figure.

Table 2. Suggestions for how to assess the elements in three different level of detail of energy audit program evaluation study.

Level of detail of evaluation study	Discount rate [%]	Free-rider effect	Data collection method	Multi-variate approach	Performance indicators			
					Firm cost effective-ness	Program cost effectiveness	Total energy savings	CO$_2$ reduction
Minimum require-ments for evaluation study	Does not have to be used in calculations	Should be accounted for	Might be included	Does not have to be used in calculations	Does not have to be calculated	Should be calculated and presented in [€/MWh]	Should be calculated	Does not have to be calculated
Recom-mended require-ments for evaluation study	At least one should be used in calculations of firm cost effectiveness	Should be accounted for	Should be included	Might be used in calculations	Should be calculated	Should be calculated and presented in [€/kWh] and [€]	Should be calculated	Might be calculated
Thorough evaluation study	Several should be used in calculations of firm cost effectiveness	Should be accounted for as a question in interview or questionnaire	Should be included	Should be used in calculations	Should be calculated and presented in [€/kWh]	Should be calculated and presented in [€/kWh] and [€]	Should be calculated	Should be calculated

7. Conclusions and policy implications

The aim of this paper was to evaluate the comparability of evaluation studies of energy audit policy programs by investigating five different energy audit policy program evaluations. The choice of method and the sample of evaluations affect the results but the conclusions would not likely change even if more evaluations had been studied. Results show that the categorization of EEMs differs between the evaluation studies, which makes them difficult to compare. As a conclusion, more consistency in how categorizations are made is needed. This could be included in available standards for energy audits, e.g. ISO 50002, and would greatly increase the possibility for a common categorization in different evaluation studies, which would increase their comparability. More comprehensive energy audits would also improve the quality of available energy data and proposed EEMs, as well as facilitate for industrial companies to work continuously with energy efficiency and would be a first step towards introducing energy management.

Performance indicators for the presented evaluation studies differ from each other. As already mentioned, this makes comparison difficult. Guidelines for what performance indicators should be presented in evaluation studies as well as how they should be calculated would increase the comparability. The guidelines should state the lifetime of EEMs, which discount rates to use, etc. However, even if standards are possible in respect of how to conduct an energy audit policy program evaluation, standards for policy research such as evaluations of energy efficiency policy programs might not be accepted by scientists from different backgrounds.

The policy implication of this paper is that new energy audit policy programs must distinguish a harmonized way of categorizing data for both measures and energy end-use. This will leave policy makers and researchers with an answer to the essential question: in which processes is energy in fact used, and where is the largest potential found? Further, a standard for how to evaluate energy audit policy programs is suggested that allows comparison between countries and regions. A harmonized data categorization and a harmonized standard for how to evaluate audit programs are not least needed within the EU, where member states are obliged to launch audit programs for large enterprises, and preferably also for SMEs.

Such a standard should include a minimum of assessments of included elements. As the free-rider effect is important as regards the cost effectiveness of a program, this should always be included as a question in an evaluation study. As for the performance indicators, the most important to be addressed in an evaluation study are program cost effectiveness and total energy savings. The calculation of these should be similar, made in a straightforward way, and presented in Euro per unit of energy end-use, preferably also using a few different discount rates but most importantly the same discount rate. The other elements in an evaluation might be considered to be assessed in a more detailed, thorough energy audit program evaluation, although still include the minimum criteria. The development of a methodology to use, however, should be further researched, where a protocol for energy audit program design should be developed as well.

Acknowledgements

References

Anderson, S.T., Newell, R.G., 2004. Information programs for technology adoption: The case of energy-efficiency audits. Resour. Energy Econ. 26 (1), 27–50.

Basurko, O.C., Gabiña, G., Uriondo, Z., 2013. Energy performance of fishing vessels and potential savings. J. Clean. Prod. 54, 30–40.

Bertoldi, P., 2001. Effective Policies and Measures in Energy Efficiency in End-Use Equipment and Industrial processes. In: Proceedings of the 2001 Workshop on Good Practices in Policies and Measurers.

EEFIG (Energy Efficiency Financial Institutions Group), 2015. Energy Efficiency - the first fuel for the EU Economy. http://ec.europa.eu/energy/sites/ener/files/documents/Final%20Report%20EEFIG%20v%209.1%2024022015%20clean%20FINAL%20sent.pdf (accessed 16.09.28).

Farla, J., Blok, K., 2002. Industrial long-term agreements on energy efficiency in the Netherlands. A critical

assessment of the monitoring methodologies and quantitative results. J. Clean. Prod. 10 (2), 165–182.

Fleiter, T., Gruber, E., Eichhammer, W., Worrell, E., 2012a. The German energy audit program for firms - a cost-effective way to improve energy efficiency? Energy Effic. 5 (4), 447–469.

Fleiter, T., Schleich, J., Ravivanpong, P., 2012b. Adoption of energy-efficiency measures in SMEs - An empirical analysis based on energy audit data from Germany. Energy Policy 51, 863–875.

Harris, J., Anderson, J., Shafron, W., 2000. Investment in energy efficiency: A survey of Australian firms. Energy Policy 28 (12), 867–876.

International Organization of Standardization, 2014. ISO 50002:2014, Energy Audits - Requirements with Guidance for Use.

Kaval, P., Baskaran, R., 2013. Key Ideas and Concepts from Economics for Understanding the Roles and Value of Ecosystem Services, in: Wratten, S., Sandhu, H., Cullen, R., Costanza, R., (Eds.) Ecosystem Services in Agricultural & Urban Landscapes. John Wiley & Sons, Inc., pp. 28-41.

Lublin, Z., Lock, A., 2013. Energikartläggningscheckar - En samhällsekonomisk utvärdering. Swedish Energy Agency, Eskilstuna. (In Swedish)

Mattsson, B., 2006. Kostnads-nyttoanalys för nybörjare. Räddningsverket, Karlstad. (In Swedish)

Petek, J., Glavič, P., Kostevšek, A., 2016. Comprehensive approach to increase energy efficiency based on versatile industrial practices. J. Clean. Prod. 112, 2813–2821.

Price, L., Lu, H., 2011. Industrial energy auditing and assessments: A survey of programs around the world. in: ECEEE Summer Study 2011. Energy Efficiency First: the Foundation of a Low-carbon Society, Proceedings, pp. 629–640.

Schulze, M., Heidenreich, S., 2016. Linking energy-related strategic flexibility and energy efficiency – The mediating role of management control systems choice. J. Clean. Prod. (In press)

Schulze, M., Nehler, H., Ottosson, M., Thollander, P., 2016. Energy management in industry – a systematic review of previous findings and an integrative conceptual framework. J. Clean. Prod. 112, 3692-3708.

Stenqvist, C., Nilsson, L.-J., 2012. Energy efficiency in energy-intensive industries - an evaluation of the Swedish voluntary agreement PFE. Energy Effic. 5 (2), 225–241.

Thollander, P., Danestig, M., Rohdin, P., 2007. Energy policies for increased industrial energy efficiency: evaluation of a local energy programme for manufacturing SMEs. Energy Policy 35 (11), 5774–5783.

Thollander, P., Dotzauer, E., 2010. An energy efficiency program for Swedish industrial small- and medium-sized enterprises. J. Clean. Prod. 18 (13), 1339–1346.

Thollander, P., Paramonova, S., Cornelis, E., Kimura, O., Trianni, A., Karlsson, M., Cagno, E., Morales, I., Jimenez Navarro, J.-P., 2015. International study on energy end-use data among industrial SMEs (small and medium-sized enterprises) and energy end-use efficiency improvement opportunities. J. Clean. Prod. 104, 282–296.

Thomas, S., Boonekamp, P., Vreuls, H., Broc, J.S., Bosseboeuf, D., Lapillonne, B., Labanca, N., 2012. How to measure the overall energy savings linked to policies and energy services at the national level? Energy Effic. 5 (1), 19–35.

Vedung, E., 1997. Public Policy and Program Evaluation. Transaction Publishers, New Brunswick.

Vine, E., 2008. Strategies and policies for improving energy efficiency programs: Closing the loop between evaluation and implementation. Energy Policy 36 (10), 3872–3881.

Vine, E., Hall, N., Keating, K.M., Kushler, M., Prahl, R., 2012. Emerging issues in the evaluation of energy-efficiency programs: The US experience. Energy Effic. 5 (1), 5–17.

Väisänen, H. (co-ordinator), Christensen, W., Despretz, H., Espegren, K.A., Gaspar, C., Lytras, K., Meyer, B., Reinikainen, E., Sattler, M., Starzer, O., 2003. Guidebook for Energy Audit Programme Developers. http://www.motiva.fi/files/1805/GB_Printversion.pdf (accessed 16.09.28).

Weiss, C.H., 1998. Evaluation: methods for studying programs and policies, second ed. Prentice Hall, Upper Saddle River, New Jersey.

Yin, R.K., 2014. Case Study Research: Design and Methods, fifth ed. SAGE, London.