

# Did you behave as we designed you to? Monitoring and evaluating behavioural change in Demand side management: from what to why

Ruth Mourik, Duneworks, Eschweileroord 57, 5625 NN Eindhoven, The Netherlands

Luc van Summeren, Duneworks, Eschweileroord 57, 5625 NN Eindhoven, The Netherlands

Sea Rotmann, SEA - Sustainable Energy Advice, 43 Moea Point Road, 6022 Wellington, New Zealand

Sylvia Breukers, Duneworks, Eschweileroord 57, 5625 NN Eindhoven, The Netherlands

## **Abstract**

One of the key contemporary challenges facing energy Demand Side Management interventions is finding the right ways to monitor and evaluate the intervention and its actual and preferably longitudinal behavioural change impacts. In this paper we will first briefly explain monitoring and evaluation, its current practice and importance to policymaking and how different disciplinary underpinnings of interventions influence this. The current focus on outputs instead of outcome in monitoring and evaluation practices of DSM interventions largely follows the economic and psychological underpinnings of most interventions in the field. However, the field increasingly develops more systemic, sociology-underpinned types of interventions, and this does require a rethinking of our current monitoring and evaluation practices. In addition, there are many challenges one currently faces when attempting to monitor and evaluate behavioural change in DSM interventions in general. These challenges lead us to conclude that the traditional quantitative proxies used at present often do not correctly reflect the real behavioural changes that occur. We reflect on how a learning process around monitoring and evaluation could look like that is relevant to end-users, 'cost effective', doable, measures actual behavioural change, focuses on both the individual and societal level, allows for different definitions of success and flexibility in changing goals and methods, and provides learning about the processes underpinning that change. We conclude with proposing an alternative which challenges common beliefs, addresses new issues, and proposes innovative methods to the current mainstream approach. This includes a focus on double-loop learning, allowing for different definitions of success and creating a more participatory approach focused on both process and outcome. In addition, this new approach makes use of a combination of qualitative and quantitative metrics to evaluate a multitude of parameters for success. This process should be entered as a collective and collaborative learning process involving policymakers, funders, researchers, end-users, technology developers and other stakeholders involved in systemic DSM interventions.

## **Introduction**

One of the key contemporary challenges facing energy Demand Side Management interventions is finding the right ways to monitor and evaluate the intervention and its actual and preferably longitudinal behavioural change impacts. In this paper we will first briefly explain monitoring and evaluation, its current practice and importance to policymaking and how different disciplinary underpinnings of interventions influence this. The current focus on outputs instead of outcome in monitoring and evaluation practices of DSM interventions largely follows the economic and psychological underpinnings of most interventions in the field. However, the field increasingly develops more systemic, sociology-underpinned types of interventions, and this does require a rethinking of our current monitoring and evaluation practices. In addition, there are many challenges one currently faces when attempting to monitor and evaluate behavioural change in DSM interventions in general. These challenges lead us to conclude that the traditional quantitative proxies used at present often do not correctly reflect the real behavioural changes that occur. We reflect on how a learning process around monitoring and evaluation could look like that is relevant to end-users, 'cost effective', doable, measures actual behavioural change, focuses on both the individual and societal level, allows for different definitions of success and flexibility in changing goals and methods, and provides learning about the processes underpinning that change. We conclude with proposing an alternative which challenges common beliefs, addresses new issues, and proposes innovative methods to the current mainstream approach. This includes a focus on double-loop learning, allowing for different definitions of success and creating a more participatory approach focused on both process and outcome. In addition, this new approach makes use of a combination of qualitative and quantitative metrics to evaluate a multitude of parameters for success. This process should be entered as a collective and collaborative learning process

involving policymakers, funders, researchers, end-users, technology developers and other stakeholders involved in systemic DSM interventions.

## **Evaluating efficiency and effectiveness of DSM interventions**

The House of Lords (2011) states that evaluation of DSM interventions is necessary to be able to assess four issues: establishing the effect of the intervention, assessing the need for improvements, assessing the value for money and contributing to the development of an evidence-base for the effectiveness of behavioural interventions at the population level. In other words: extrapolating the evidence of effectiveness on individual or household level to the general population. In many conventional DSM interventions, efficiency and effectiveness are used as indicators for success or failure of interventions in reaching pre-set goals. Effectiveness refers to the extent in which an intervention reaches the intended goals. Goals to be evaluated usually consist of: a reduction in energy consumption, energy savings, number of homes retrofitted etc., but sometimes benefits are included that fit in a broader energy context, e.g. health improvements, job creation and safety improvements (Mourik & Rotmann, 2013).

In addition, some (Breukers et al. 2009) emphasise that particularly when it concerns behavioural change, another element that needs to be taken into account when evaluating effectiveness is the lasting effect beyond the duration of an intervention. This applies to both habitual and one-off or one-shot decisions. We differentiate between one-shot decisions that are performed rarely and consciously e.g. investing in energy efficiency improvements and habitual behaviour which is more frequent and less conscious, e.g. showering, changing the settings of the thermostat etc. In terms of energy-using behaviours most (up to 95%) of our behaviours are thought to be habitual (Darnton et al, 2011) even though one-off investment decisions, like insulating a home, can lead to greater, immediate energy savings. Lasting changes in habitual behaviour, however, will continuously lead to ongoing energy savings. According to Breukers et al (2009), in this definition of effectiveness, an energy DSM intervention is highly effective when it has reached its goals and/or has had a positive effect on reducing total energy consumption and when it has led to lasting behavioural change and energy savings in the target group. Evaluating this lasting parameter of effectiveness is, however, a major challenge, as will be discussed in the next section.

Efficiency is usually measured in terms of cost-effectiveness, which compares the inputs and outputs of a DSM intervention. These cost-effectiveness calculations can be made from various perspectives (e.g. participants, service providers, intermediaries and society). Often, cost-effectiveness is measured by the Net Present Value of the impacts of a intervention; this NPV is calculated by dividing the sum of benefits by the sum of the costs of the intervention (Breukers et al 2009). For more information about basic concepts, calculation rules and systems for Energy Savings Calculations see Task 21 of the IEA DSM IA (Vreuls et al 2012). In addition, more information about evaluation with a focus on outcomes and energy impacts can be found in the Evaluation Guidebook of IEA DSM IA (Vreuls et al 2012).

Monitoring and evaluating the efficiency of an intervention is considered important because funders often have an interest in assessing how effectively the money was spent. In addition, Rosenow and Galvin (2013) state that resources have to be allocated among competing energy efficiency interventions and effectiveness is believed to be a good indicator of how money should be spent. The IEA (2014) demonstrates that, consequently, many benefits are usually monetised so that they can be integrated into existing policy assessment frameworks. This is one reason for many evaluators to prefer quantitative over qualitative data and focus on the monitoring of quantifiable elements of the intervention.

The above focus on quantitative forms of effectiveness and efficiency in monitoring and evaluation fall under what is called single-loop learning (Argyris & Schön 1978). Single-loop learning is about the effectiveness and/or efficiency of a technology, measure, instrument, arrangement, or intervention to achieve pre-defined goals. Single-loop learning involves mainly instrumental learning about the (often short-term) effectiveness and efficiency of measures in reaching the goals. This type of learning sees behavioural change interventions more or less as linear cause and effect relationships ( $A+B=C$ : Intervention A targeted on group B will cause the intended Change C) (Shove 2010). The above can also be characterised as the difference between a focus on output and outcome, where outputs are direct and measurable products of an intervention, whereas outcomes refer to the results and impact and or improvements in the short, medium and long-term. In short, the difference being made by an intervention. The Logic Model workbook ( [www.innonet.org](http://www.innonet.org) p17) uses the following citation to illustrate the difference between outputs and outcome:

*“Not how many worms the bird feeds its’ young, but how well the fledgling flies.” –  
(United Way of America, 1999).*

With this quote in our minds, we can argue in favour for an approach that addresses the ‘fledglings’ rather than the ‘worms’. However, in practice we rarely witness evaluation focused on outcomes. The sections below will discuss this emphasis on single-loop learning and explore how double-loop evaluations could look like and point towards potential ways forward for improving monitoring and evaluating practices in the field of energy DSM. However, before doing so, we will first address very broadly how different disciplinary perspectives propose to monitor and evaluate. A detailed discussion on different disciplinary models of understanding behaviour and theories of changing it can be found in Mourik and Rotmann (2013) Subtask 1 Deliverable (‘the Monster’) or on our [Task 24 wiki](#)<sup>1</sup>.

## **Disciplinary basis for interventions and consequences for monitoring and evaluating behavioural change**

DSM interventions are often (mostly implicitly) based upon economic, psychological and sometimes on sociological models of behaviour or behaviour change. Kok et al (2011) argue that within those disciplines the economic and psychological approaches have been dominant, e.g. in many countries policymakers have favoured DSM interventions based more upon economic than psychological theory. The more sociological approaches to energy DSM interventions are only recently emerging (Breukers et al 2009).

It is useful to consider the disciplinary underpinnings of interventions since each discipline has its own focus and units of analysis and the different disciplinary underpinnings of interventions affect goals and thus influence what can and should be monitored. In this paper we focus on the monitoring and evaluation approaches that follow from the disciplinary foci discussed in more detail in the final Subtask 3 Deliverable 3A of IEA DSM Task 24. We realise that we are very categorical and that we do not do proper justice to the diversity within and cross-overs between disciplines, but it is not our aim to be comprehensive.

Economic-, psychological- and/or sociological models propose different views on how and to what extent energy behaviour can be influenced. None of the disciplines claim to provide a complete picture of energy behaviour, but all to a certain extent simplify the complex issues surrounding energy behaviour to make them better manageable. In practice, Chatterton (2011) argues, approaches often use a mix of disciplines and disciplines themselves also overlap sometimes on certain aspects, e.g. behavioural economics, which combines concepts of economic and psychological theories. Both economic and psychological approaches see energy-related behaviour as a cause-effect relationship. Psychology does appreciate the role of context (the facilitating or hindering role of the environment). In addition, psychology crucially focuses on peoples' subjective perceptions and constructions of the world, e.g. even if everything is ‘objectively’ the same (demographics, technology available), what a person thinks about energy saving (their attitudes, perceptions, values) plays an important role. Energy DSM interventions underpinned by this line of thinking usually target the individual and follow a barrier/driver approach. In addition, monitoring and evaluation of these type of interventions often focus on single-loop learning, on outputs and consequently, also on cost efficiency and other quantitative parameters and less so on actual behaviour change. Sociological approaches see energy-related behaviour as more complex; often accept its habitual nature, and increasingly view energy demand as a practice which is dependent on many contextual factors. Interventions underpinned by this disciplinary approach usually take a more systemic approach. These systemic interventions require more systemic evaluation methods that allow an understanding of the interdependent but not cause-effect type of relationship and mechanisms and attempt to develop indicators for behavioural change that allow for more realistic evaluation focused on outputs and outcomes.

Both DSM interventions and their evaluations are diverse, are based (explicitly or implicitly) on diverse disciplines and may target various kinds of problems and contexts. There is no silver bullet or best model or combination of models that can be used for designing, implementing or evaluating all types of DSM interventions. And criticising an economic- or psychology-based intervention for its lack of monitoring of social practices is not doing justice to the discipline underpinning the intervention. Different approaches will be effective for targeting different problems and reaching different goals. Different types of behaviour might also best be targeted by using different approaches, but this requires further research. Furthermore, different audience groups will respond differently to different kinds of interventions. In a later section we will discuss the current policy practice and the space, interest and also fear of monitoring and evaluation present in that arena.

---

<sup>1</sup> <http://www.leadstask24wiki.info>

When faced with complex societal problems, policymakers and practitioners often reach for single-loop type of projects that simplify targeted problems, e.g. reducing the questions of energy demand to individuals' choices and using proxies such as saved kWh or number of people reached with a campaign to determine its effectiveness (see also Mourik and Rotmann, 2013). Darnton and Evans (2013) state that such a reduced problem definition informs many DSM interventions and is fine when one deals with an intervention only aimed at informing people about possible renovation choices, or when an intervention is aimed at installing as many technological energy efficient solutions as possible. However, when the aim is to realise lasting behavioural/habitual changes, increasingly, voices (Gynther et al 2012; Lai Fong et al. 2014) are stating and demonstrating that multidisciplinary approaches in general tend to be more effective than approaches derived from only one discipline. In addition, the current output-focused practice of monitoring and evaluation is increasingly criticised for its short-termism and voices argue for more outcome-focus in monitoring and evaluation of energy efficiency and DSM interventions.

The discussion of monitoring and evaluation and the influence of different disciplines also exemplifies that the current focus on outputs instead of outcomes in monitoring and evaluation practices of DSM interventions largely follows the economic and psychological underpinnings of most interventions in the field. However, the field increasingly develops more systemic, sociology-underpinned types of interventions, and this does require a rethinking of our current monitoring and evaluation practices. In the next section we will discuss these more systemic interventions and the monitoring and evaluation issues accompanying them in more detail, but start first with a discussion of challenges the current monitoring and evaluation practices face when asked to determine the behavioural change impact of interventions.

### **The multiple challenges of monitoring and evaluating behavioural change in DSM interventions**

Monitoring and evaluating DSM interventions is clearly a complex field, with no clear answers yet to several pertinent questions and challenges. Below we briefly list several of these interlinked challenges and observations from the field that are pertinent to monitoring and evaluating energy DSM interventions. For a more extensive discussion we refer to the underlying report delivered under subtask 3 of Task 24.

#### ***A lack of benchmarking, which is an adequate instrument to measure improvements against a set base-line***

Obviously, benchmarking is not straightforward or easy on a lot of occasions - the historical data about energy consumption of a household is often not easily accessible, or not even relevant when the intervention targets households that moved to a newly built neighbourhood. If the benchmark makes use of best practice interventions, then it is very difficult to design exactly the same intervention or it may not even be pertinent if the context is different. And benchmarking becomes even more of a challenge when trying to benchmark elusive things such as practices, habits and behaviours.

#### ***Focus is often mostly at the implementation stage (supply side) instead of the in-use phase (end users), which means that the occurrence and evolution of behavioural change is not addressed***

Monitoring and evaluation is currently often aimed at the implementation stage, not the actual use phase. This follows from the fact that monitoring and evaluating is aimed at reflecting upon issues relevant to the designers and the funders, not the customers or end-users or any intermediaries involved in the roll-out of the intervention (e.g. insulation installers). To put it strongly, for a smart meter rollout intervention, in the end the number of units installed will define whether or not the intervention was a success, irrespective of the actual savings achieved through the smart meter or if it created e.g. a healthy market or provided customers with better quality information (Batey, Mourik and Garcia 2014). M&E should aim to understand the how, how long, who, where, when and why of an intervention and thus much more explicitly focus on the customer and end-user benefits and needs and monitor and evaluate these (House of Lords 2011).

#### ***In cases where behavioural change is addressed, there is often a lack of longitudinal M&E, which makes it impossible to assess the long-term outcome***

Monitoring and evaluation often takes place in a once-off manner, focusing on the implementation phase or just after. Longitudinal monitoring and evaluation does not fit the attitudes in the public sector (nor the private sector) that are often short-term and focused on cost-effectiveness. Funding for longitudinal evaluation is, consequently, often not available. However, if the aim of monitoring and evaluation is also on learning how long the effect of an intervention persists, and if it is at least sustained and not undone through rebound effects or

because of people reverting to their earlier behaviour, longitudinal monitoring and evaluation is necessary. In addition, if evaluation is geared at one specific type of energy-related behaviour, it might also be that rebound effects remain undetected

***The M&E team is often only involved after the intervention is concluded. Consequently, requirements for monitoring and evaluation are often not included in the fine-tuning of an intervention***

This might follow from the fact that it is often unclear in the design phase which results will demonstrate success or failure (House of Lords 2011). The National Centre for Sustainability (2011) however, argues that monitoring and evaluation should be planned during the design phase. Thinking about monitoring and evaluation can even help to clarify what the aims and goals of an intervention are. This brings us to another challenge, namely that most monitoring and evaluation is not explicitly part of a DSM intervention when it is actually being implemented. Many interventions only perform monitoring and evaluation after an intervention has ended. One reason for this is the fact that behavioural interventions are often publicly funded and therefore, it is regarded as important that they provide value for the invested money. This can be evaluated simply by collecting numbers and other quantitative information, which is best done right after the intervention (House of Lords 2011). Learning about what works, when, where, for who, how (long) and why is not really the focus of many interventions

***Monitoring is often based on modelling and irrelevant proxies of behaviour, like for instance energy savings, cost savings, number of homes retrofitted and the floor area insulated etc.***

Fell and King (2012, pp33) describe this challenge as follows: "Understanding the how and why people use gas and electricity as they do is key to designing interventions that successfully increase energy efficiency." Designing research projects that are aimed at doing exactly this, through working directly with users to co-design interventions and their evaluation, is a small but growing field (Lockton et al, 2013). Karlin et al (in prep) demonstrate that there is no standard way of collecting data and assessing what actual changes in behaviour may (or may not) have led to observed kWh changes. Monitoring and evaluation of the impact of an intervention on behaviour is often determined by means of modelling outputs such as energy savings, cost savings, number of homes retrofitted, floor area insulated etc. If a focus on behaviour is attempted, this usually involves a focus on identifying behavioural determinants, or the use of self-reported behaviour or willingness-to-pay records. Actual observation and or metering of behaviour is hardly undertaken. Certainly not on a large scale, at best with a small sample. In practice people usually make more complex trade-offs between costs and gains (both financial and non- financial) than assumed in models and, consequently, the models are most likely not accurate and estimated savings in evaluations are often higher than actual savings.

***The distribution of costs and benefits between different stakeholder is often not monitored and evaluated, while this is crucial to understand why end-users have responded the ways they did***

In general only costs on the supply side are calculated. This method does not take into account costs that are incurred by the end-users at whom the intervention is targeted, e.g. time spent making a decision, cleaning up a loft before insulation, costs of broken belongings or repairs needed after an intervention, soft costs such as dealing with negative social feedback, or the costs of changing habitual behaviour (it may be more inconvenient or take a little more time). It is indeed difficult to put a number on these costs, but it can be done (Breukers et al. 2014). Another issue related to the above is that there is an uneven distribution of costs and benefits in many DSM interventions, where the burden is usually laid on the end-users. For example, when installing insulation or EE appliances and installations in a home it should not be left to the individual to buy and install metering devices to be able to meter the actual impact of retrofitting.

***Different stakeholders may have different definitions of success, which are often not made explicit. In addition to this, e.g. end-users' success definitions are often not identified, monitored or evaluated at all***

The generally used cost-effectiveness may not capture many of the potential social welfare outcomes and/or impacts such as job creation, positive health effects, reduced environmental externalities etc. Moreover, interventions may have positive spill-over effects that not only influence the target group (e.g. neighbouring effect) but have larger systemic impact and longer-term effects. In addition, energy consumers often value other features besides cost reductions which are not included in these cost-benefit calculations (e.g. health, safety, quality improvements). This demonstrates that potentially, evaluating successfulness of an intervention should allow the identification of multiple definitions of success.

## **Focus is often on individuals and not on practices or socially-shared ways-of-doing among social groups**

### ***In traditional M&E there are no participatory elements or feedback loops***

Current monitoring and evaluation practices are usually rather top-down exercises, not involving the end-users. Or, if they do involve them, it is at most in the form of user-generated data such as self-reported accounts around attitudes and willingness (Karlin, in prep). Often, end-users do not even receive feedback on the impact of the intervention, certainly not on a collective level, but sometimes not even at the level of the individual household. And if the household wants to monitor and evaluate the impact it is sometimes forced to buy their own M&E technologies, e.g. displays. However, when designing an intervention that also accounts for end-user success indicators, that allows for an understanding of the for who, when, where, how, how long and why of behavioural change, and that accounts for the distribution of hard and soft costs and benefits. Involving end-users explicitly is a logical step considering they are potentially best positioned to understand the context in which they operate (Lockton et al. 2013).

### ***Conventional measurements of success may not capture many of the potential additional or multiple benefits of an intervention (e.g. health, comfort, convenience)***

The currently-used proxies and methodologies are not only limited to provide the right information to assess the cost-effectiveness of interventions, but that the focus on cost-effectiveness itself fails to provide a good evaluation of the impact of an intervention on behavioural change and even more so, on lifestyle changes. In addition, the proxies and/or indicators used today are often not able to capture the long-term outcomes nor outcomes that go beyond the targeted behaviour or targeted individual such as e.g. the creation of a market, or increased competences of tradespeople, or increased health are not evaluated, whilst these actually allow for a potential upscaling of interventions in the absence of government intervention. In the Warm Up New Zealand case study, for example, it was found that the wider macro-economic health benefits of the national insulation subsidy were \$5 to every \$1 spent (Mourik and Rotmann, 2013)! This good news story would have never seen the light of day if an evaluation team that was strongly focused on public health outcomes hadn't been chosen.

With the above challenges in mind we argue for a practice of monitoring and evaluating that allows for involvement of more stakeholders, including end-users, and allows for the measuring of both outputs and outcomes, and of both energy-related and other behaviours, and multiple impacts in the short, medium and long term. Or in other words, we argue for monitoring and evaluation that allows for an understanding of the when, where, for whom, how, how long and especially why an intervention was or wasn't effective in changing energy behaviour. In the sections below we will discuss the type of monitoring and evaluation that might provide a more 'realist' approach and one that encourages widespread learning.

## **Opening up interventions to include end users and what makes an intervention successful to them, and what that means for M&E**

There are multiple ways of undertaking evaluation and monitoring. Very roughly speaking, approaches move from those that are narrow, concrete and short term in focus towards those that take a broader, more systemic and longer-term focus. We have illustrated that the choice for an intervention in theory also sets the parameters for monitoring and evaluation in practice. As discussed above, current practice is to favour DSM interventions based upon economic or psychological theories. But multidisciplinary (including sociological) approaches with the following characteristics are emerging: they are more tailored, multidisciplinary, consist of varied interventions, use both qualitative and quantitative indicators, are iterative and flexible, have a systemic approach, are multi stakeholder, often more participatory and have outcomes beyond the duration of an intervention and beyond energy. In addition, there is evidence that a more participatory policy process has a positive effect on actual energy savings (Batey, Mourik and Garcia 2014; Pianosi, Bull and Rieser 2012). Increasingly, voices are stating and demonstrating that multidisciplinary approaches to behaviour change in general tend to be more effective in realising lasting changes than approaches derived from only one discipline (Gynther et al. 2012; Lai Fong et al. 2014).

We do appreciate the enormous challenge of this change in focus in interventions for their monitoring and evaluation, the potential costs involved, the complexity of managing such flexible goals, multifaceted, interdependent, contextualised, multi-output and outcome and multi-stakeholder processes. In addition, each policy evaluation may turn out to have impact on the policymaking process if the conclusion is that money was badly spent. The political reality is that policymakers need to be able to explain how their policy (positively) impacts a large group, not only a very small one. Therefore, policymakers need information that is scalable to

the national level in a quantified form. We need to acknowledge this mandate and the complexity of policymaking and policy decision-making processes. Consequently, it is very understandable that policymakers demand simple, focused, and quantitative up-scaled evaluations defining success in efficiency and effectiveness terms (Batey, Mourik and Garcia 2014; Shove 2010, Broenner and Moller 2013). A change of focus amongst policymakers and funders towards allowing experimentation with more systemic and messy real life interventions that do not provide easily quantifiable and scalable information is a big transition. It demands amongst others that policymakers appreciate that these systemic interventions cannot be evaluated in terms of cause and effect, but are the outcome of a complex process (one where they can only play a finite and relatively small role in). Different methods and approaches to monitoring and evaluation will be used in practice and the policy context will determine the exact nature of these - but we strongly recommend/encourage the use of more systems-based, learning-focused monitoring and evaluation regimes. Embarking on such a learning process and could demand more radical DSM policymaking (Shove in Task 24 Oxford Workshop<sup>2</sup>). This process should thus be entered as a collective and collaborative learning process involving policymakers, funders, researchers, end-users, technology developers and other stakeholders involved in systemic DSM interventions.

But how could a learning process around monitoring and evaluation look like that is relevant to end-users, 'cost effective', doable, measures actual behavioural change, focuses on both the individual and societal level, allows for different definitions of success and flexibility in changing goals and methods, and provides learning about the processes underpinning that change? Most methods available now unfortunately still very much rely on systematic reviews that are less suitable to understand the why and how of complex interventions, such as the successful comprehensive energy DSM intervention [Sustainable Järva](#)<sup>3</sup> (Pawson et al. 2005; Mourik and Rotmann 2013). To understand the how and why monitoring and evaluation needs to account for effects of context and (place in) time. And understanding context is important if we want to learn how to make the results up-scalable under different circumstances.

### Embarking on a collective learning process: How and what can we learn from monitoring and evaluating a systemic intervention?

As discussed above, in order to learn more about DSM interventions and to improve current and future interventions it is important to learn about the processes leading to behavioural change and how to monitor and evaluate them. We propose to focus on the concept of double-loop learning as a potential framework for this learning process. The single- and double-loop learning concepts were developed by Argyris and Schön (1978). The figure below highlights the complementarity between single- and double-loop learning.

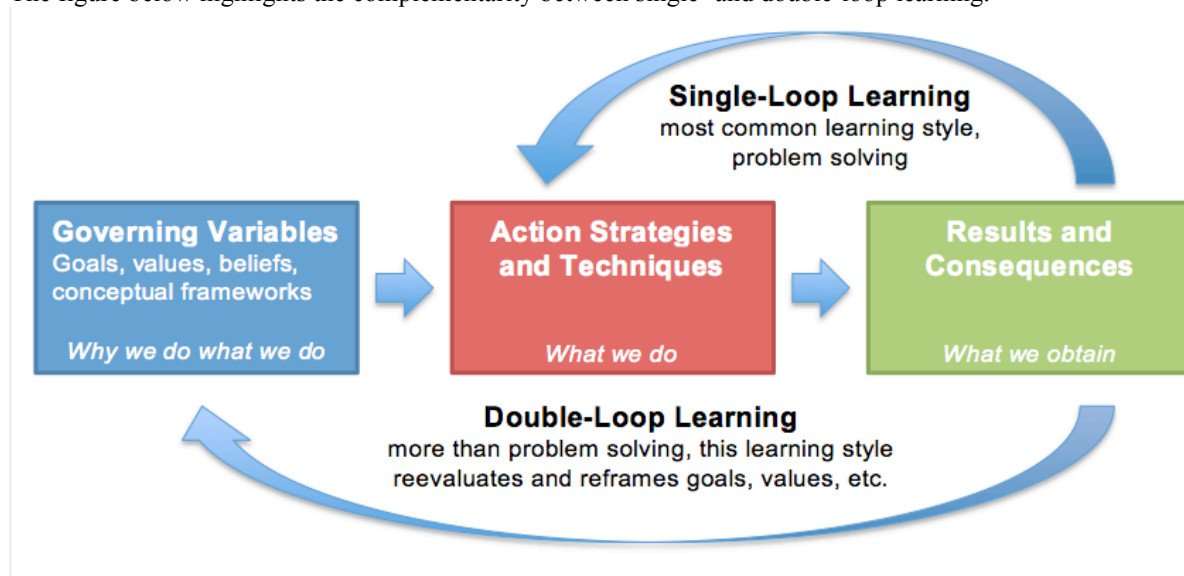


Figure1: double vs single loop learning. Retrieved from <http://www.afs.org/blog/icl/?p=2653>

To summarise, in the context of DSM interventions, single-loop learning is instrumental and focused on short-term learning about effectiveness in meeting goals. It is outcome focused. Double-loop learning includes learning about single-loop learning and takes an additional step and involves time and context and reflection and

<sup>2</sup> [http://www.leadstask24wiki.info/wiki/Subtask\\_1](http://www.leadstask24wiki.info/wiki/Subtask_1)

<sup>3</sup> [http://www.leadstask24wiki.info/wiki/Sustainable\\_Järva](http://www.leadstask24wiki.info/wiki/Sustainable_Järva)

focuses on both strategic and operational results. Double-loop learning is process-oriented, focused on the how, when, where, how, how long, for whom and is about questioning goals and the prevailing norms and rules underlying these goals. In addition, double-loop learning is focused on interactions, the quality of participation, learning by doing and doing by learning, aligning expectations, in short, double-loop learning is about reflexive governance of interventions (Breukers et al., 2009). Breukers et al. (2009) state that this type of learning may lead to changes in the way the intervention manager learns in interaction with stakeholders (e.g. end-users) and this learning can change both the contents and context of the intervention. It will change the way how stakeholders frame problems, find solutions and manage and understand their own role. Double-loop learning is seen as a process in which learning is an important precondition for systematic transitions to take place. Indicators that focus on double-loop learning can be used to evaluate DSM interventions and to see whether they contribute to long-term, broader and more lasting changes (Breukers et al. 2009). In the table below, single- and double-loop learning, their main indicators and questions that can be answered by monitoring these indicators are shown. These questions will be used in Deliverable 3B to provide more concrete guidelines for using single- and double-loop learning in M&E.

Table 3: Indicators for evaluating successful learning processes (adapted from Breukers et al, 2009)

| Learning type        | Indicators  | Questions for M&E   | Metrics (examples)   |
|----------------------|---|---|--|
| Single-loop learning | Efficiency indicators:<br>Cost-effectiveness<br>Goals reached (within given time and allocated budget)  | Was the intervention cost-effective?<br>Are the goals reached within the time and within the allocated budget?  | Costs and benefits (eg RoI or NPV)<br>Pre-set goals<br>Available time and time needed<br>Budget and costs  |
|                      | Effectiveness indicators:<br>Reaching the intended goals<br>Lowering the total energy consumption   | Are the goals reached?<br>Is the total energy consumption lowered (per household? by sector?)   | Energy savings<br>Energy consumption before and after intervention   |
| Double-loop learning | Process indicators:<br>Realising a network of a heterogeneous set of actors with different definitions of success<br>Interaction and participation by the target group (so that they can learn about their own behaviour and consequences for energy consumption)<br>Interaction and participation with a diverse set of stakeholders since the design phase<br>Learning as an explicit aim of the intervention<br>Record new lessons for future interventions<br>Making use of lessons that are learned during previous interventions<br>Perspectives of intermediaries before and after a intervention<br>Changes in assumptions, norms and beliefs | To what extent is a network of a heterogeneous set of actors developed in which they all participated and interacted with each other since the design phase? Did this lead to different definitions of success?<br>How was interaction and participation by the target group allowed in the programme? And to what extent did end-users learn about their own behaviour and consequences for their energy consumption?<br>How was learning during and after the intervention ensured?<br>How did the perspectives, assumptions, norms and beliefs of intermediaries and other stakeholders change during the programme? | Diversity of actors that are involved in the design and implementation of the intervention<br>Definitions of success that were co-created and used<br>The way end-users were involved in the design and implementation of the intervention<br>Perceived self-efficacy<br>Perceived impact and benefit of the intervention<br><br>Learning strategy<br><br>Perspectives, assumptions, norms and beliefs of stakeholders before, during and after the intervention |



|  |  |   |  |
|--|--|---|--|
|  | <p>Content indicators:</p> <p>Alignment of the expectations of the stakeholders</p> <p>Reflection upon the function of evaluation/monitoring together with stakeholders</p> <p>Learned lessons during the intervention are translated into (re)designs</p> <p>Improving the capacity of own or similar organisations to perform successful DSM interventions</p> <p>Creation of new networks and institutions that support the newly formed behaviour and its outcomes</p> <p>Lasting changes (behavioural or practice change)</p> | <p>To what extent were the expectations of stakeholders aligned? How is this done?</p> <p>How did reflection upon the function of M&amp;E with stakeholders take place?</p> <p>Which lessons learned during the intervention are translated into (re)designs?</p> <p>Is the capacity of own- or similar organisations improved to perform successful DSM interventions?</p> <p>Are new networks and institutions created that support the newly formed behaviour and its outcomes?</p> <p>Did lasting changes take place?</p> | <p>Collective impact approach to co-develop metrics to measure this</p> <p>Main lessons learned by different stakeholders</p> <p>Perceived success of collaboration and intervention design and implementation</p> <p>Short- and long-term effects</p> |
|--|--|---|--|

**Concluding remarks**

The current monitoring and evaluation practice is one where decision-makers and funders often choose to support interventions on the basis of the possibility of evaluating these with existing methods. These interventions are evaluated with single-loop learning indicators, output indicators, and insufficient focus is put on outcome, both short-term and long-term. Interventions are evaluated too often on the basis of modelled projections of savings and improvements and too little on the basis of real measurements. A problem with this situation is that it actually recreates and sustains the status quo in interventions and monitoring and evaluation and does not elicit actual double-loop learning.

We have argued that this monitoring and evaluation practice stands in stark contrast with what we should desire instead: getting stakeholders to start reflecting on the issue of proxies and monitoring and evaluating successful, ongoing behaviour change. There are many ways to measure, but not everything that can be measured matters and not everything that matters can be measured. The question is if we need to try to define new indicators to monitor and evaluate the systemic interventions? Do we need to define with which tools or methods such systemic interventions can be monitored and evaluated? Do we need to engage with scaling up towards identifying indicators that actually represent behaviour change and are thus better proxies? Do we need to identify indicators that assess the outcomes of behaviour change and not so much focus on behaviour change itself? Or are we too far removed from that ideal scenario, and will identifying more indicators and working with these create a false sense of reassurance?

So instead of continuing funding and supporting only those projects that are aimed at establishing the single-loop learning types of outputs and that can be evaluated by means of the current proxies and methods we might also chose to explicitly engage more with projects where we explicitly target to collectively and collaboratively learn (also about learning) about behaviour change.

A focus on learning also implies involving more different types of stakeholders in the design, implementation and monitoring and evaluation of these projects and approaching all of them with the clear aim to learn from them. Learning thus implies not only to learn within the research arena but also within the policy arena and funding arena. We are at a crossroad and need to embark on a learning process, the researchers, intermediaries, end-users and policymakers alike and collaboratively.

This will be necessary in order to learn how to better design policy, research and interventions around monitoring and evaluating more comprehensive interventions with an explicit aim to generate cross-cutting impacts relevant to different stakeholders. These stakeholders often do not work in the same department, organisation or sector and therefore, this focus on multiple benefits and impacts also implies shared learning and the creation of a shared language and framework across sectors and departmental silos that each hold a piece of the puzzle.

It is very important to map the different stakeholders that are imperative to the successful conclusion of an intervention and to understand their unique mandates, needs and restrictions and any points of conflict that can arise between them. This is all part of the current energy system, which operates in a largely neoliberal political and economic setting.

An effective way to also report on the learning process is to focus explicitly on these learning stories which are in essence a process of co-design and dialogue and retrace replicable elements in these learning stories to allow for a more successful delivery of comprehensive energy efficiency DSM interventions (Moezzi and Janda 2014). Storytelling is an effective dialogue and evaluation tool, it allows for multiple perspectives and creates a deeper appreciation for the fact that there is not one truth. It allows to move beyond the presented and pretended objectivity of a more quantitative approach. It not only allows for different morals to be discussed, it almost demands it, we are all aware of the almost inherited right of stories to have multiple interpretations depending on the reader, so instead of either accepting or opposing a story, readers are encouraged to try to understand a story and its multiple interpretations. Through the telling of stories the listeners and presenters learn, not only about negative or unintended consequences. But they also learn to experience bad experiences as part of learning and turning points in a story, with the aim to do better next time. This approach could be transformational but only if the actors are willing to participate in the process. There is need for some pre-conditions to exist in the workplace for staff to believe that there is something to gain by telling their story. These same pre-conditions may not be as important at home. We therefore need to investigate these different conditions and contexts.

Analysing and evaluating the different stories could be undertaken by means of realist synthesis as developed by Pawson et al. (2005): “Realist review is a relatively new strategy for synthesising research, which has an explanatory rather than judgemental focus. Specifically, it seeks to ‘unpack the mechanism’ of how complex programmes work (or why they fail) in particular contexts and settings” (Pawson, et al. 2005). Pawson’s approach allows for history and place, i.e. path dependency to be taken into account.

In a parallel movement we of course need to work on identifying and quantifying relevant indicators for different kinds of stakeholders and for different types of behaviour to allow for a mix of qualitative and quantitative monitoring and evaluation. One interesting approach might be to learn from the measuring of behavioural change in other areas such as health and education and translate these to the energy efficiency field. Political voting and viewer ratings could be interesting examples to learn from as well. Another important step we need to take is to provide a clear overview of different energy management practices in homes and the potential impact they might have and design indicators to measure and understand the why and how for each and every one of these different types of behaviours (Fell and King 2012). And more work is currently conducted on this topic, see for example Karlin and Ford (in prep) who will design a tool consisting of both qualitative and quantitative indicators to evaluate behavioural change (see also Subtask 9 of the Task 24 extension). See also Ucci et al. (2013) who developed a novel questionnaire-based benchmarking tool for the assessment of behaviour change potential for energy saving in office and industrial settings.

To conclude, we recommend a more negotiable and flexible practice of monitoring with a mix of both quantitative and qualitative indicators. With a view to further narrowing the field of enquiry a large sample can then be monitored by means of quantitative indicators and then the intervention needs becoming smart about identifying users to work with, and approach these selected users with more qualitative methods to understand the, where, when, whom, how and why. An interesting practical example is discussed in the study produced by Fell and King (2012) of Brook Lyndhurst for DECC which followed up quantitative sampling with qualitative interviews/home visits to understand why comparable households use different amounts of energy. The study demonstrated how quantitative modelling was able to explain less than 40% of the variation in gas consumption of similar households. To understand the variation qualitative research was undertaken to explore day-to-day lives of 70 high and low gas user (50/50%) households. Methods used were interviews, house tours, diary exercises and unobtrusive temperature monitoring. The analysis led to a valuable clustering of three behaviour types (temperature management, people in the home and physical properties of the home) that explained the variations.

Finally, we do recommend a more decentralised collective participatory approach to monitoring and evaluation, allowing for the involvement of all relevant stakeholders, but end-users in particular (Batey et al forthcoming 2015). Co-design of interventions should become mainstream practice. There are democratic issues with the current design of interventions, where top-down approaches with interventions designed by a select group are common practice and a more democratic approach, including more co-design would be beneficial. End-users are uniquely positioned to understand the when, where, who, how and why interventions succeed or fail and are potentially best equipped to understand their place in time and context and its influence on the intervention's outcome (Lockton et al. 2013). This co-design or more participatory approach would also provide means to initiate a discussion regarding the privacy issues related to monitoring and evaluating behaviour change by external parties (Darby 2010, Hargreaves, Nye & Burgess 2010, Breukers & Mourik 2013). Additionally, the ICT revolution and smart meter roll-out in Europe which have the potential to empower end-users to start engaging with do-it-yourself monitoring, or user generated content, are important next research fields (Batey, Mourik and Garcia 2014). And in doing further research on this topic we need to keep in mind what Sarah Darby said in the closing panel of the [2014 Behave conference](#)<sup>4</sup>: we need "metrics of resilience" that also reflect the need to keep our eye on the future as designers and evaluators.

We are aware that this paper is more a positioning paper than a useful guide for policymakers. This report might even add more to the complexity of evaluation in the policy domain than simplify it. We have proposed to use a definition of 'energy behaviour' to include a wide range of energy-related actions/decisions that range from large and "once in a lifetime" investments to the routines of everyday living. A fundamental premise for Task 24 is that the techno-economical paradigm that used to (and to some extent still does) dominate this field should be challenged by alternative models and theories. In particular it is argued that practice theory or at least sociological approaches are potent "challengers".

We are explicitly not stating that the alternative theories should replace the traditional theories aimed at understanding energy behaviour. They should complement them and fill gaps we are experiencing in understanding the complexity of everyday life and the energy use and needs accompanying these complexities. In addition, as Hal Wilhite discussed in his closing keynote at the 2014 Behave conference in Oxford, we are facing a societal transition in terms of how we produce and use and distribute energy and we need to understand the energy effects of that transition. And the traditional approaches taking a basically economic approach to behaviour (and as such are mainly useful to understand one-shot behaviour and less so for habitual behaviour) are simply not sufficient to tackle that challenge. Vice versa, the sociological approaches might be less suited to understanding one-shot behaviour.

The same mechanism indeed also applies for the monitoring and evaluation indicators and tools linked to the different disciplines. The calculation and modelling type of monitoring and evaluation and accompanying single-loop learning process are very well suited to assess the success of subsidy schemes or installation schemes. However, to understand complex multifaceted interventions, sociological tools and metrics focused on double-loop learning are necessary.

For a more detailed discussion on the appropriateness of different theories to understand different types of behaviour we refer to Deliverable 1 of Task 24: the 'Monster' report and [Wiki](#), where we discuss different theories and approaches in detail, including their effectiveness to explain or influence specific types of behaviour. Finally, this paper is a summary of several reports delivered under Subtask 3 of IEA DSM Task 24.

## References

Argyris, C., & Schön, D. (1978) *Organizational learning: A theory of action perspective*, Reading, Mass: Addison Wesley.

Batey, M., Mourik, R.M., J.P. Garcia (2014) From calculated to real energy savings performance evaluation: how ICT-innovation and Do-It-yourself user-generated data monitoring can enable to improve real performance evaluation of energy-efficiency initiatives. Paper presented at the IEPPEC conference, Berlin, 2014.

---

<sup>4</sup> <http://behaveconference.com/>

Breukers, S., Backhaus, J., Mourik, R., Hodson, M., Marvin, S., & Brohmann, B. (2010). Practicing Learning and Learning in Practice; Testing learning tools for energy demand side management projects. Knowledge Collaboration & Learning for Sustainable Innovation.

Breukers, S. C., Heiskanen, E., Brohmann, B., Mourik, R. M., & Feenstra, C. F. (2011). Connecting research to practice to improve energy demand-side management (DSM). *Energy* 36, 2176-2185.

Breukers, S., Heiskanen, E., Mourik, R. M., Bauknecht, D., Hodson, M., Barabanova, Y., Vadovics, E. (2009). Interaction Schemes for Successful Energy Demand Side Management. Building Blocks for a practicable and conceptual framework. *Changing Behaviour*.

Breukers, S.; Mourik, R.M. (2013). End-user engagement for flexible energy consumption patterns. Towards tailored interventions for shifting and reducing household energy demand. paper presented at InContext Scientific Workshop 7-8 October 2013, Delft, The Netherlands.

Breukers, S, Mourik, R. M., Backhaus, J., Mathijssen, T, Brunsting, S, Uytterlinde, M, & Pol, M. (2013). [Effectief beleid voor duurzaam gedrag: Een internationale vergelijking. Studie voor de Raad voor de Leefomgeving en Infrastructuur \(RLI\)](#). (p. 105). Raad voor de Leefomgeving en Infrastructuur (RLI).

Breukers, S.; Summeren, L. van, Mourik, R.M. (2014). Eerst proces, dan prestatie. Naar een optimale afstemming tussen aanbieders, bewoners en de woning in renovaties richting energieneutraliteit. Hoofdrapportage in opdracht van Platform 31.

Chatterton, T. (2011). An Introduction to Thinking About 'Energy Behaviour': a multi-model approach. Bristol: Department of energy & climate change.

Darby S (2006) Social learning and public policy: lessons from an energy-conscious village. *Energy Policy* 34, 2929-2940

Darby S (2010) Smart metering: what potential for householder engagement? *Building Research and Information* 38 (5), 442-457

Darby, S. (2010). Literature review for the energy Demand Project. Environmental Change Institute, University of Oxford.

Darnton, A., & Evans, D. (2013). *Influencing Behaviours; A technical guide to the ISM tool*. Edinburgh: The Scottish Government.

Darnton, A, Verplanken, B, White, P and Whitmarsh, L (2011). *Habits, Routines and Sustainable Lifestyles: A summary report to the Department for Environment, Food and Rural Affairs. AD Research & Analysis for Defra, London*.

Fell D., King G. (2012). Domestic energy use study: to understand why comparable households use different amounts of energy A report to the Department for Energy and Climate Change. Brook Lyndhurst. DECC, London.

Gynther, L., Mikkonen, I., & Smits, A. (2012). Evaluation of European energy behavioural change programmes. *Energy Efficiency*, 67-82.

Hargreaves, T., M. Nye and J. Burgess. 2010. "Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors". *Energy Policy* 38 (2010): 6111-6119.

House of Lords. (2011). *Behaviour Change*. Retrieved from <http://www.parliament.uk/business/committees/committees-a-z/lords-select/science-and-technology-committee/inquiries/parliament-2010/behaviour/>

IEA. (2014). *Capturing the Multiple Benefits of Energy Efficiency*. International Energy Agency.

Karlin, B., R. Ford et al (in prep) Beyond kWh: Methodology Review for Assessing Behavior-Based Energy Interventions

Kok, G., Hing Lo, S., Peters, G.-J. Y., & Ruiter, R. A. (2011). Changing energy-related behavior: An Intervention Mapping approach. *Energy Policy*, 5280-5286.

Lockton, D., Harrison, D., & Stanton, N. A. (2012). Models of the user: designers' perspectives on influencing sustainable behaviour. *Journal of Design Research*, 7-27.

Lockton, D., Bowden, F., Greene, C., Brass, C. & Gheerawo, R. (2013). 'People and energy: A design-led approach to understanding everyday energy use behaviour'. EPIC 2013: Ethnographic Praxis in Industry Conference, 15-18 September 2013, Royal Institution, London. (doi: 10.1111/j.1559-8918.2013.00029.x)

Moezzi, M., & Janda, K. B. (2014). From "if only" to "social potential" in schemes to reduce building energy use. *Energy Research & Social Science*, 30-40.

Mourik, R.M. and S. Rotmann. Most of the Time what we do is what we do most of the time. And sometimes we do something new. Analysis of Case-studies IEA DSM Task 24 Closing the Loop – Behaviour Change in DSM: from Theory to Practice. Deliverable 2 for IEA Implementing Agreement DSM Task 24. September 2013.

Pawson, R., Greenhalgh, T., Harvey, G., Walshe, K. (2005). Realist review--a new method of systematic review designed for complex policy interventions. *Journal of Health Serv Res Policy*. 2005 Jul;10 Suppl 1:21-34.

Pianosi, M., Bull, R., and Rieser, M. (2012) Enhancing Environmental Citizenship and Reducing Energy Consumption through Creative Engagement with Building Users. Conference Proceedings for 2012 International Energy Program Evaluation Conference, Rome, Italy.

Rosenow, J., & Galvin, R. (2013). Evaluating the evaluations: Evidence from energy efficiency programmes in Germany and the UK. *Energy and Buildings*, 450-458.

Shove E, 2010, "Beyond the ABC: climate change policy and theories of social change" *Environment and Planning A* 42(6) 1273 – 1285

Ucci, M. (2013) Behaviour change potential for energy saving in non-domestic buildings: Development and pilot-testing of a benchmarking tool. *Building Services Engineering Research & Technology*, vol. 35, no. 1. pp 36-52.

Vreuls, H. (2005). Evaluating energy efficiency policy measures & DSM programmes. International Energy Agency. Retrieved from <http://www.ieadsm.org/Files/EXCO%20File%20Library/Key%20Publications/Volume1Total.pdf>

Vreuls, H., & Both, D. (2012). Harmonised Energy Savings Calculations for selected end-use technologies, key elements and practical formulas. IEA. Retrieved from IEA DSM: <http://www.ieadsm.org/Files/Tasks/Task%2021%20-%20Standardisation%20of%20Energy%20Savings%20Calculations/final%20public%20version/Report%20on%20Energy%20savings%20calculation%20final%20version.pdf>