International Energy Agency

Energy Technology Initiative on

Demand Side Management Technologies and Programmes

Did you behave as we designed you to?

Task 24 – Phase I Subtask 3

Closing the Loop – Behaviour Change in DSM:
From Theory to Practice

Deliverable 3A – Positioning paper on monitoring & evaluation

February 2015

R.M Mourik, L.F.M. van Summeren,
S. Breukers, S. Rotmann

With contributions from, in alphabetical order: Tim Chatterton, Kady Cowan, Sarah Darby, Svetlana Gross, Sarah Higginson, Katy Janda, Dan Lockton, Erica Löfström, Malcolm McCulloch, Corinna Moser, Sabine Pahl, and Janet Stephenson

Contents

Prologue

The three Subtask 3 deliverables at quick glance

When we developed the work plan for Task 24 one of the starting points was the appreciation that DSM projects demonstrate great diversity in goals, scope, participants, resources etc to match the diversity of Behaviour Changers' contexts and needs and their wider environment. As a consequence, developing a generic evaluation and monitoring framework that is widely applicable, yet does justice to this diversity, is very difficult indeed. We identified that finding more appropriate, effective and maybe also standard ways of monitoring, evaluation and learning about successful DSM implementations was a real and urgent need. Currently, DSM policymakers and other relevant Behaviour Changers usually fund and/or support DSM programmes on a rather ad-hoc basis because they lack these means of assessing their impact on contributing towards a more sustainable energy system.

We felt that a review of state of the art research findings and current best practice and potential standardised ways of monitoring and evaluating could identify what roles and actions policymakers, investors and other Behaviour Changers might play to make behaviour change for DSM successful when tapping into the vast and cost-effective potential for energy efficiency and conservation. This review was undertaken under the umbrella of the Task by Karlin et al, 2015 (Deliverable 3). Although this Deliverable will become fully available in the course of the extension of Task 24 (Subtask 9), Karlin et al already were able to find that there is no standardised way of monitoring the impact of behavioural change DSM interventions beyond kWh type of indicators (and often even they are not measured in a standardised way). One of the consequences is that research funders lack clear evaluation frameworks of successful outcomes to decide on funding practical behaviour change research efforts and thus continue relying on the 'easier', technological fixes to our energy problems and the more common economic or psychological theory underpinned type of interventions. The more complex systemic type of interventions that go beyond mere kWh type of outputs thus face severe start-up issues.

Building on a large database, the conclusion of this Methodological Review (Deliverable 3) was that there was a dire need for a wider discussion about these complex issues. A discussion on how a more standardised, practical, robust generic evaluation and monitoring framework to evaluate both kWh type of outputs as well as longer term behavioural outcomes contributing to a more energy efficient DSM system would look like. We provided a first attempt at initiating and contributing to such a discussion with our second deliverable, this positioning paper (Deliverable 3A).

However, our Task also promised it would create an internationally-validated monitoring and evaluation template that would contribute to the ability to monitor, evaluate and prove ongoing success of behaviour change outcomes leading to energy and CO_2 savings, health and social benefits, financial savings and community engagement. In our third Deliverable (3B), the factsheet document, we attempted to develop a practical, context-specific monitoring and evaluation template for various DSM tools (which can be used alone or in combination in behavioural interventions), with the specific aim to meet various Behaviour Changers' needs for outcome evaluation. This template is developed to match the monitoring and evaluation analysis in Subtasks 1 and 2 of Task 24. The factsheets are a template (completed for 3 types of intervention tools in the Building Retrofit domain: Energy Performance Certificates, mass marketing campaigns and subsidy schemes) which aims at providing indicators, metrics and ways to monitor and evaluate long-term, identifiable and/or measurable behaviour change outcomes of DSM programmes. These indicators aim to be context-sensitive and contingent on the sector/goals/target groups of behaviour change interventions.

In our extension (Subtasks 8 and 9) we will further build on these factsheets, develop them for each of the domains and all the known tools used in interventions. In addition, the 'Beyond kWh' tool (Subtask 9) will provide clear insights into what data will need to be collected depending on the interven- tion/tools used, and in what way. As such, the last Deliverable (3B) is a living document.

Introduction

One of the key contemporary challenges facing energy Demand Side Management (DSM) interventions is finding the right ways to monitor and evaluate an intervention and its actual, and preferably, longitudinal behavioural change impact.[1]

In this report[2] we will first briefly explain what monitoring and evaluation mean, its current practice and how different disciplinary underpinnings of behaviour change interventions ('interventions' from now on) influence this. We will then continue discussing the 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 (which are often collected ad hoc and in a non-standard way, see Karlin et al) often do not correctly reflect the details of the real behavioural changes that occur. Solely quantitative assessment often misses the details of what exactly is going on, for different people (end users and stakeholders) and in different contexts. This is problematic for multiple reasons, and we conclude with proposing an alternative to the current mainstream approach. This alternative 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 that makes use of a combination of qualitative and quantitative metrics to evaluate a multitude of parameters for success.

Definitions

In this Task 24, we have clearly defined some of the major definitions we use in our reports:

Energy behaviour refers to all human actions that affect the way that fuels (electricity, gas, petroleum, coal, etc) are used to achieve desired services, including the acquisition or disposal of energy-related technologies and materials, the ways in which these are used, and the mental processes that relate to these actions.

Behaviour Change in the context of this Task thus refers to any changes in said human actions which were directly or indirectly influenced by a variety of interventions (e.g. legislation, regulation, incentives, subsidies, information campaigns, peer pressure etc.) aimed at fulfilling specific behaviour change outcomes. These outcomes can include any changes in energy efficiency, total energy consumption, energy technology uptake or demand management but should be identified and specified by the Behaviour Changer designing the intervention for the purpose of outcome evaluation.

Behaviour Changer is a person or agency tasked with the goal of designing, implementing, evaluating and/or disseminating interventions geared at changing energy end users behaviours. In this Task, we differentiate between five Behaviour Changer sectors: 'the Decisionmaker' (Government on all levels), 'the Provider' (Energy-providing industry on all levels including technology manufacturers), 'the Expert' (researchers and consultants from a multitude of disciplines especially economics, psychology, sociology and engineering), 'the Conscience' (the Third sector including NGOs, community organisations, transition towns etc.) and 'the Doer' (the intermediaries selling energy-using goods and services who are directly in contact with the end users).

Demand Side Management are interventions (top-down and bottom-up policies, programmes and actions) developed and performed by Behaviour Changers (e.g. government

[1] Mourik, R., & Rotmann, S. (2012). *IEA DSM TASK XXIV: Closing the Loop - Behaviour Change in DSM: From theory to practice*; Task Work Plan. Wellington, NZ; Eindhoven, NL: IEA DSM Implementing Agreement.

[2] The report is part of one of three deliverables for Subtask 3 which aim to focus on the development of context sensitive indicators, metrics and ways to monitor and evaluate long-term, identifiable and/or measurable behaviour change outcomes of DSM interventions within the four themes of this task (building retrofits, transport, SME's and smart meters/feedback). A methodological review of Task 24 case studies and other literature on the behavioural data collected in feedback and building retrofit interventions in the residential sector has been undertaken by Karlin et al and forms the background to Subtask 3. This report presents the general observations and lessons learnt based on a literature review on evaluation, the empirical analysis of cases submitted to Task 24 and discussions and workshops with national and external experts during the Task workshops in the UK, September 2014 and Graz, October 2014. In the third part factsheets will be presented for one of the four themes of the Task with concrete guidelines for monitoring and evaluation of different behaviour change interventions.

agencies, utilities, DSM implementers) that seek to influence the ways in which end users consume energy at home, at their workplace or whilst travelling. The changes sought by Behaviour Changers may include the quantity of energy consumed for a given service, patterns of energy consumption or the supply management and type of energy consumed.

Evaluation is a structured process of assessing the success of an intervention in meeting its goals and to reflect on the lessons learned during the intervention.

Monitoring refers to measuring progress and achievement, and whether the planned outputs and outcomes are produced.

Effectiveness refers to the extent in which an intervention reaches the intended goals.

Efficiency is usually measured in terms of cost-effectiveness, which compares the inputs and outputs of a DSM intervention.

One-shot or investment behaviours are performed rarely and consciously e.g. investing in energy efficiency improvements.

Habitual behaviour is more frequent and less conscious, e.g. showering, changing the settings of the thermostat etc.

Outputs are direct and measurable products of an intervention.

Outcomes refer to the results and impact and or improvements in the short, medium and long-term.

Single-loop learning is about the effectiveness and/or efficiency of a technology, measure, instrument, arrangement, or intervention to achieve pre-defined goals.

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.

A benchmark is a point of reference from which measurements may be made.

A DSM tool is used in DSM interventions to pursue a desired outcome (e.g. lowering or shifting energy demand). Examples of DSM tools are subsidies, mass media campaigns, energy labelling, regulation etc.

Introducing monitoring and evaluating behavioural change in DSM interventions[3]

In this report **monitoring** refers to *measuring progress and achievement, and whether the planned outputs and outcomes are produced*. **Evaluation** is a *structured process of assessing the success of an intervention in meeting its goals and to reflect on the lessons learned during the intervention.* Evaluating and monitoring have a clear function, e.g. additional learning to enhance future interventions or re-iterate current ones, giving account to funders of past interventions etc. Depending on the function of the intervention, different tools and approaches are selected for monitoring and evaluating.

According to the National Centre for Sustainability (2011), the main difference between monitoring and evaluation is that evaluation explicitly places a value judgement on the information and data monitored and gathered during the intervention. However, what to monitor is of course a normative judgement as well, since it defines what would count as success parameters, what questions will be answered and which data will be collected. Often, trade-offs have to be made between the quality and resources available for the monitoring and evaluation, and this may affect the choice for different indicators and metrics.

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:

1. establishing the effect of the intervention,
2. assessing the need for improvements,
3. assessing the value for money and
4. contributing to the development of an evidence-base for the effectiveness of behavioural interven- tions 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[4].

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

[3] We are using the term intervention to refer to a project, programme or policy. We are not going into potential differences between projects, programmes or policies regarding their M&E.

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

retrofitted etc., but sometimes benefits are included that fit in a broader energy context, e.g. health improvements, job creation and safety improvements.

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. See both figures below for an overview of the types of behaviour interventions can target and the differences between these behaviours[5] :

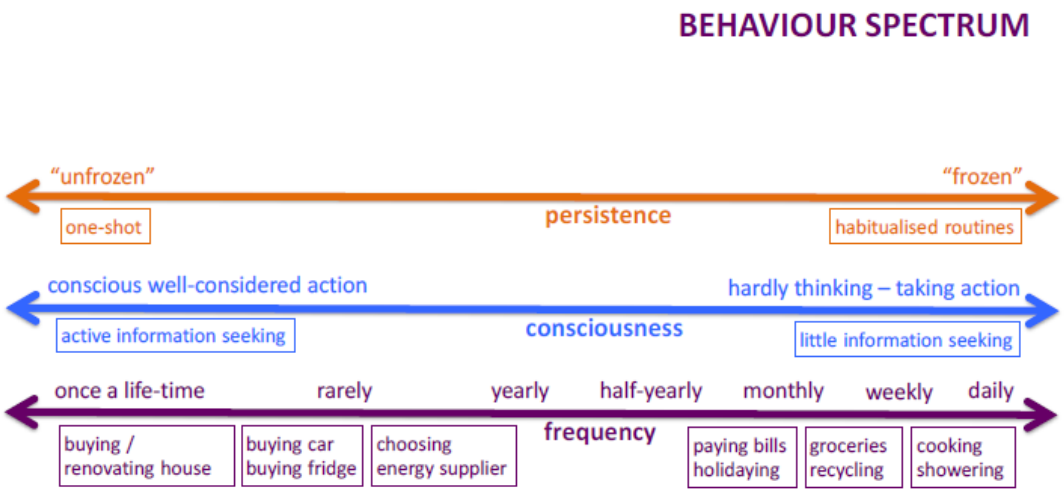


Figure 1: Behaviour spectrum, retrieved from Breukers & Mourik 2013

	Green behavior Do new behavior, one that is unfamiliar	**Blue** behavior Do familiar behavior	**Purple** behavior Increase behavior intensity or duration	**Gray** behavior Decrease behavior intensity or duration	**Black** behavior Stop doing a behavior
Dot behavior is done one-time	**GreenDot** Do new behavior one time *Install solar panels on house*	**BlueDot** Do familiar behavior one time *Tell a friend about eco-friendly soap*	**PurpleDot** Increase behavior one time *Plant more trees & local plants today*	**GrayDot** Decrease behavior one time *Buy fewer bottles of water now*	**BlackDot** Stop doing a behavior one time *Turn off space heater for tonight*
Span behavior has specific duration, such as 40 days	**GreenSpan** Do new behavior for a period of time *Carpool to work for three weeks*	**BlueSpan** Do familiar behavior for a period of time *Bike to work for two months*	**PurpleSpan** Increase behavior for a period of time *Take public bus for one month*	**GraySpan** Decrease behavior for a period of time *Take shorter showers this week*	**BlackSpan** Stop a behavior for a period of time *Don't water lawn during summer*
Path behavior is done from now on, a permanent change	**GreenPath** Do new behavior from now on *Start growing own vegetables*	**BluePath** Do familiar behavior from now on *Turn off lights when leaving room*	**PurplePath** Increase behavior from now on *Purchase more local produce*	**GrayPath** Decrease behavior from now on *Eat less meat from now on*	**BlackPath** Stop a behavior from now on *Never litter again*

Figure 2: Foggs' behaviour grid of 15 types of behaviour change

[5] Some other possibly useful reference here are frameworks/spectrums/classifications developed by Gary Raw et al. (2010) (in a domestic energy context).

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.[6] For more information about basic concepts, calculation rules and systems for Energy Savings Calculations see Task 21 of the IEA DSM IA.[7] 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.[8]

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.[9] 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[10]. The Logic Model workbook[11] uses the following citation to illustrate the difference between outputs and outcome:

---

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

[7] 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

[8] 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

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

[10] Taken from the "Outcomes vs. Outputs" section of Innovation Network's Logic Model Workbook (p. 17). The entire workbook is available after free registration on www.innonet.org (last accessed December 10, 2008), and from PowerPoint presentation slides created by Taylor-Powell, E., & Henert, E. (2008). Developing a logic model. Retrieved January 2009, from University of Wisconsin-Extension-Cooperative Extension, Program Development and Evaluation Unit website at http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html#more

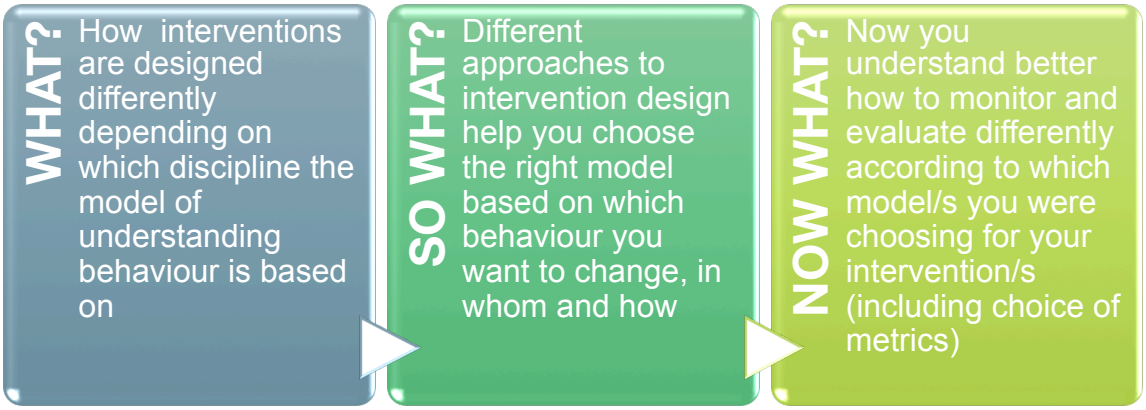
[11] Idem

"*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.

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.[12] 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.

Three disciplinary perspectives used in DSM programmes are discussed (economics, psychology and sociology). We do this in order to assess what makes a 'successful' intervention according to these perspectives and what this implies for evaluation. 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.

The aim of this chapter is to exemplify the main differences between approaches based upon these disciplines. We have chosen to follow the division proposed by Darby (2010) and Chatterton (2011)

[12] 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.

complemented with viewpoints by Breukers et al. (2011), Darnton & Evans (2013) , and Darnton et al. (2011). In this chapter we aim to merely illustrate how evaluation and monitoring approaches relate to the initial conceptual choices made in the design of an intervention.

Economic theories

Chatterton (2011) summarises the view of economic theories on energy related behaviour as follows: "*Energy is a commodity and consumers will adapt their usage in response to price signals*" (Chatterton, 2011, p31).

Darnton et al (2013) state that in the **neo-classical economic** perspective behaviours are seen as (semi-) rational decisions that are made through cost-benefit calculations. Several market failures are seen as the main barriers for energy efficiency; these market failures consist of a lack of information on the risks and benefits of new solutions and lack of access to capital for investments (Breukers et al. 2011). In this perspective an individual energy user is seen as someone who makes his or her energy-related choices in a rational/semi-rational manner, and consequently, the behaviour is assumed to be fairly predictable.

These economic or techno-economic approaches have been criticised by Breukers et al. (2009) because they focus only on limited aspects of behaviour in practice, and underestimate the social meanings and contexts of energy usage, which are important in realising behavioural changes. This led to many developments in the field of economic theories concerning (energy) behaviour, appropriating psychological theory elements, such as the emergence of **behavioural economics** and **Nudge Theory** that amongst others make use of the notion of bounded rationality as introduced by Herbert Simon. "*Bounded rationality asserts that decision makers are intendedly rational; that is, they are goal-oriented and adaptive, but because of human cognitive and emotional architecture, they sometimes fail, occasionally in important decisions*" (Jones, 1999, p297). With their 'ecological rationality' perspective Gerd Gigerenzer and Peter Todd (2012) argue that Herbert Simon's introduction of the term bounded rationality to economics was not a criticism of human 'failure' in decision making, but a recognition that people making everyday decisions under constraints (bounds) of time/knowledge/willpower will - quite rationally given the constraints - simplify decision processes and use heuristics such as choosing the first option that works, or establishing 'sub-optimal' habits because they get the job done. When you apply it to energy use, you can understand why routines like turning the thermostat all the way up, or closing the blinds and switching the lights on, are 'rational' given the constraints of everyday life - they work, and achieve the contextual goal people are after (warmth, consistent light, etc.) even if in the long term they are not 'rational'.

Behavioural economics and nudging combine economic theories with theories and concepts of psychology that do grasp social meanings and other relevant contexts of energy usage more effectively.

Psychological theories

According to Darby (2010) and Chatterton (2011) many psychological approaches view energy related behaviour as follows: "*Energy use can be affected by stimulus – response mechanisms and by engaging attention*" (Chatterton, 2011, p32). Stimulus-response was an approach in the 1960s, before the 'cognitive revolution' and since then psychology also deals with affect and emotion, motivation, values, thinking and reasoning, interpersonal and group processes etc.

Just as in economic theories, the individual takes a central role in psychological research. However, here are different branches in psychology, and for example, in **social psychology**, the individual is always seen as embedded in groups and therefore, processes of cognition (e.g., perception, categorisation), and behaviour and decisions are not seen as purely individualistic but embedded in and influenced by their social context. It is thus increasingly acknowledged that this individual also operates as part of a collective e.g. by imitating the behaviour of important others, or through the influence of social norms, social comparison, social learning.

Many psychological approaches view decision-making of an individual as a mental calculation aimed at making choices; these calculations are informed by both emotion and cold calculus. Darnton et al.

(2011) conclude that behaviour is often seen as the product of someone's beliefs, attitudes and other motivational factors.

According to Kok et al (2011) there is consensus among experts in the psychological discipline that people will change their deliberate behaviour when there is a strong positive intention (attitude, perceived social norms, behaviour is consistent with self-image, positive affective reaction and perceived self-efficacy), no environmental constraints and when they have the skills necessary to perform the behaviour.

Darby (2010), Chatterton (2011) and Lockton et al. (2012) state that in many psychological theories it is also acknowledged that people's behaviour is influenced by context, however behaviour is still seen as fairly predictable and manageable. Interventions based on psychological disciplines are often designed around the concept of barriers and drivers for the desired behaviour, and these are in turn often derived from attitudinal survey data. The intervention is, consequently, built around these barriers and drivers (Darnton & Evans 2013). Another element of critique (Gynther et al. 2012), (Kok et al. 2011) is that many interventions based on psychological disciplines are directed at deliberate (rational) behaviour, whilst over ninety per cent of energy-related human behaviour is impulsive or habitual.

This view of consumption as an individual decision-making process (based in economic and psychological approaches) has been dominant in the sustainable energy and environmental policy arenas. Sahakian and Wilhite (2014) state that, consequently, interventions have focussed mainly on prompting individual behavioural change and designing more energy efficient technologies, e.g. feedback devices. However, to allow for a wider appreciation of the social context and the role of the collective in behavioural change, increasingly, the design of interventions starts to include elements of sociological theories (implicitly or explicitly).[13]

Sociological theories

Chatterton (2011) states that sociological theories look at energy-related behaviour in the following way: "*Modern energy use is largely invisible, energy systems are complex, and daily practices are significant*" (Chatterton, 2011, p31). Moezzi and Janda (2014) state that sociological theories thus put more emphasis on the importance of the social nature of energy use and to the abilities of people to participate in change in ways that fit their own contexts and concerns. And Darnton and Evans (2013) argue that in sociological approaches the central focus is on social practices, i.e. individuals move into the background. **Social practices** refer to everyday practices; and many practices can be seen as routinised activities and habits (e.g. switching off the lights when leaving a room). Chatterton (2011) uses the following explanation of these practices: "*Practices exist beyond specific performances, they consist of interconnected sets of norms, conventions, understandings, embodied know-how, states of emotion, and arrays of material things*" (Chatterton, 2011, p23). Strengers (2011) and Hazas et al. (2012) point out that the economic assumptions that perspectives seeing individual householders as 'micro-resource managers' able to make 'constant and active choices' about their energy use, are massively over-simplistic when looking at how practices really happen in the home. Family members and other social considerations make it such that much 'behaviour' is made up of evolving, negotiated (often implicitly) norms and routines.

One emerging branch of sociology is **Practice Theory**: "*this approach shifts the analytical focus away from atomized products, technologies and individuals, towards an understanding of everyday practices, many of which include routinized activities. (...) the formation of practices draws on knowledge that is distributed between people (bodies and minds), things and culturally grounded social structures*" (Sahakian & Wilhite, 2014, p26). Practices are made of materials (objects, hard infrastructure), competences (skills and know-how) and images (meanings, ideas and interpretations). Darnton et al. (2011) argue that these three elements are not determining behavioural outcomes, instead, the coming together of these elements leads to the emergence of a certain practice.

[13] See for example Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive and reconstructive power of social norms. Psychological Science, 18(5), 429–434. doi:10.1111/j.1467-9280.2007.01917.x

Habitual behaviour or routines are mostly performed automatically without weighing up the pros and cons.[14] One important characteristic of this type of behaviour is that it usually does not change, at least not permanently, unless the external circumstances also change in such a way that it 'breaks the loop'. Interventions can therefore aim to also change the external circumstances so that people are forced to renew reasoning behind this behaviour (Gynther et al. 2012). Within sociological approaches, the main factors that are believed to influence domestic energy consumption are physical systems and infrastructures, social norms, comfort preferences, daily routines, practices and options for control, and many of these factors are difficult to monitor and evaluate, let alone their influence on behaviour or behaviour change (Chatterton 2011). A possible criticism of the sociological approach is that it may underestimate the potential for change, as when individuals decide not to conform to social norms and introduce new configurations of technology and routines.

An analysis by Breukers et al. (2010) demonstrated that realising lasting energy-behavioural changes is very difficult. Breukers et al (2009, p115) argue that "*existing behaviours are not isolated but embedded in broader socio-technical systems and strongly shaped by existing infrastructures, conventions and social structures*". In order to become durable, behavioural changes have to fit within the specific physical, institutional and social environment. Therefore, Breukers et al. (2010) argue that it is crucial to gain understanding about how behaviours are embedded in broader socio-technical systems and how these systems can be changed. In order to support lasting behavioural change on the individual level, systematic changes are needed which consist of changes in things like infrastructures, conventions, frames of thinking and social structures. Breukers et al. (2013) distinguished four elements of behavioural change which can be targeted by an intervention, shown in the table below. Monitoring and evaluation practice should focus on all these elements in order to capture the way in which an intervention aims to realise behaviour change.

Element of behaviour	Description
Individual behaviour	The way in which an intervention aims to change individual behaviour
Social norm	The way in which the behavioural change is stimulated by changing existing social norms or by using existing pro-environmental norms
Policy- and institutional context	Policies, partnerships and institutional settings which may hinder, stimulate or may be needed in order to achieve behaviour change
Physical environment	Changing or using physical elements (e.g. physical infrastructure, built environment, technology, choice environment) to enable or stimulate more sustainable behaviour

Table 1: Elements of behaviour change (Breukers et al., 2013)

Monitoring and evaluating in the different disciplines

In this section we focus on the monitoring and evaluation approaches that follow from the disciplinary foci discussed above. To summarise very briefly: both economic and psychological approaches see energy-related behaviour as a cause-effect relationship. Psychology does appreciates 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.[15]

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**.

[14] See Bas Verplanken who has worked extensively on habits from a psychological perspective.
[15] Comment from Sabine Pahl: check the 'complex interventions' framework in medicine where practitioners face similar issues - health behaviours are determined by a myriad of individual, social and societal factors

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. We will discuss this challenge in a later section.

Intervention goals and evaluation methodologies commonly used in interventions underpinned by the three disciplines discussed above are shown in the table below (this is not an extensive list, it is aimed at highlighting foci and differences).

	Goals	Methodologies	Remarks (e.g. about causal relationships)
Economic models	Outputs Cost-efficiency and effectiveness Units, and proxies e.g. number of participants, homes insulated, technologies installed, kWh saved etc. Labels	Modelling Surveys Experiments Randomised control trials	Presence of cause-effect relationship Aim is to meet a priori set goals M&E often only for duration of implementation, not longer term
Psycholo-gical models	Outputs Cost-efficiency and effectiveness Behavioural changes and analysis of respective cognitive and social processes	Surveys Self-reported behavioural changes Structured interviews Randomised control trials Surveys to identify behavioural determinants like motivations, attitudes, etc. Experiments (one of the classical psychological research methods)	Cause-effect relationships: Effect on individuals of a particular incentive, via e.g. awareness, attitude, behaviour Interfering variables like social context often not taken into account
Sociolo-gical models	Outputs and Outcomes Cost-efficiency and effectiveness Learning about what works, when, where, who, how (long) and why Learning about interdependencies Learning about co-shaping and reshaping	User accounts Time diaries Cultural probes In-depth open interviews Analysis of fit of interventions with daily life Measuring real, not modelled energy consumption Co-design	Context & mechanism/conditions produce an outcome. Direct cause-effect relationships hard to establish because of interdependencies that cannot be analysed separately

Table 2: Disciplinary perspectives, goals and evaluation metrics. From Breukers et al. (2009)

The above discussion illustrates that 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[16]. 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.

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 discuss several of these interlinked challenges and observations from the field that are pertinent to monitoring and evaluating energy DSM interventions.

[16] There are several studies by Wes Schultz and colleagues that focus on social norms. The influence of close others' attitudes is also an explicit part of the **Theory of Planned Behaviour**

Benchmarking: How do you know the intervention led to the desired effects?

A benchmark is: "*a point of reference from which measurements may be made*"[17]. This point of reference can be the historical energy consumption figures of a household or company, but it can also be best practices in DSM interventions.

When using metered outputs and especially when focusing on outcomes, benchmarking should be part of the intervention. As for metered outputs, these offer the opportunity to establish quantified achievements in terms of energy saving. However, more common is the use of random control samples, where the control group is as similar as possible (socio-demographics, type of building etc.) to the pilot or intervention participants, but where the intervention is not being applied. As such, this control group can be used to determine the influence of the intervention compared to a situation without the intervention.[18] 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.

Mismatch between the needs of stakeholders responsible for delivery of a project and those it is aimed for: Are you measuring outputs or outcomes and for whom?

A second challenge is that 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 assess outputs such as units, e.g. numbers of insulated homes, people reached by a campaign or modelled energy savings achieved during the intervention, focusing on the implementation is sufficient. 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).

If monitoring and evaluation are mainly aimed at calculating cost-effectiveness of a programme, this implies that only the needs of intervention designers and funders are taken into account. However, monitoring and evaluation should not only focus on the use period after the implementation, but very much on the actual energy use, not just the modelled use, as will be discussed below. Consequently, 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). Some researchers have started focusing on developing smart meters and in-home displays that monitor this actual time of use (Kobus et al. 2013; Van Dam et al. 2012).

M&E team is often not included in design and often not even part of intervention until after it is concluded: When should we best start with M&E?

[17] http://www.merriam-webster.com/dictionary/benchmark
[18] http://www.biology-online.org/dictionary/Control_group

Current practice is that many interventions do not invite the evaluation team to participate in the design stage. 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. In some cases, indicators or information to be collected during monitoring and evaluation can be defined at the outset. In other cases, where the design unfolds over time, perhaps in interaction with participants or target group members, this may be less the case but then it is still useful to reflect during the intervention process on how it is proceeding. However, it is also important to recognise the potential value of a more external evaluation team, which can be much more independent and objective than an internal team. An internal evaluation team might, in turn, experience more trust from the rest of the team. There is thus no clear benefit to either an internal or external team. In addition, the growing methodology of **action research** is potentially very valuable, where the researcher is independent but can feed back any issue that is monitored on a regular basis.

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, only if, in this intervention, right here, right now, it worked. In order to learn about what works when, where, for whom etc. so that an intervention can be replicated, i.e. to learn about the outcome (the bird flying), and how it came about, not just the outputs (number of worms fed), the entire process over a period of time should be evaluated, not just a snapshot at the end of the intervention (National Centre for Sustainability 2011). If the aim is only to do a one-off evaluation of the intervention without aiming to replicate this intervention or learning from it, then focusing on outputs might be sufficient. Another important reason why evaluation might be undervalued is that there are often no negative consequences when targets are not reached (e.g. How many policymakers have actually lost their jobs when an intervention failed to deliver the modelled impacts?).

Another common practice is that evaluations tend to focus on evaluating those who participated, not those who declined or opted out. Although difficult to implement in practice, it is a missed opportunity not to learn about motives for not participating in, or even opposing an intervention as this might contribute to a better embedding of future rounds of an intervention, including an iteration of better intervention design (the intervention circle of: design-implement-evaluate-iterate (and disseminate)).

Often monitoring and evaluation is not undertaken at all and behavioural interventions usually do not reserve sufficient funds for monitoring and evaluation, which should be at least 10-15% of the total intervention costs. The absence of this funding often negatively influences the quality of the evaluation (House of Lords 2011).

And a final issue on this topic is that our statement that evaluation needs to be included in design presumes that interventions have a planning phase. In practice, however, interventions sometimes are developed and implemented simultaneously. A monitoring and evaluation approach that acknowledges the lack of time allotted to up front for roll out could be very helpful.

In summary: If you want to get the best out of a programme, evaluation is essential to make it better in the future. Evaluating can allow to look back and to learn for the future.

No longitudinal M&E, sustainability/rebound often not assessed: What are the long-term effects of an intervention?

As touched upon above, 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 (e.g., if I have a more energy-efficient car, I might spend more money on holiday flights ➔ if questions only center on car use, this rebound will not be identified). This also calls for a more **systemic approach**. Adequate funding is thus needed for rigorous evaluation, especially using **population-representative samples and long-term, objective outcome measures**. In addition, investing sufficient funds into evaluation is likely to result in better value-for-money in the long-term since lessons derived from interventions can be used to improve current and future DSM interventions (House of Lords 2011).

Monitoring based on modelling and proxies: What does the data tell you about actual behavioural change?

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) have undertaken a methodological review as part of Subtask 3. It assessed all relevant Task 24 case studies and other literature in the feedback and building retrofit domains. What became quickly apparent was that there was no standard way of collecting data and assessing what actual changes in behaviour may (or may not) have led to observed kWh changes. Sometimes, even the kWh changes before and after an intervention were not measured but inferred.

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.

Explanation for this focus on modelling instead of metering actual behaviour and behavioural changes is that metering is difficult, complex in terms of privacy issues, is time consuming and costly. In addition, the time frame of both public and private sector investment is short-term, and the result is that actual change or incremental changes are not evaluated, but only modelled (Broennum and Moller 2013). When monitoring and evaluation focuses on outputs that are modelled or calculated, the costly and time-consuming process of benchmarking is not necessary. The lack of budget or willingness to allocate sufficient budget, thus might also explain the focus on modelling instead of metered data (Batey, Mourik and Garcia 2013).

One could argue that for investment behaviour the approach above might suffice. After all, if it can be deduced how many homes have been retrofitted, or how many solar panels have been installed, it is sufficient. Or is it?

There are several problems with this monitoring and evaluation of behaviour through modelling and (often quantitative) proxies. A first issue is that in the models and calculations, very often users are represented as rational actors, who will indeed act if financial gains are available. See Figure 3 based on the ABC model from Shove (2010) where she criticises the usual model of social change depending on values and attitude (A) that drive behaviour (B) where people choose (C) to adopt that behaviour. This model results in policy mainly aimed at identifying determinants of pro-environmental behaviour.

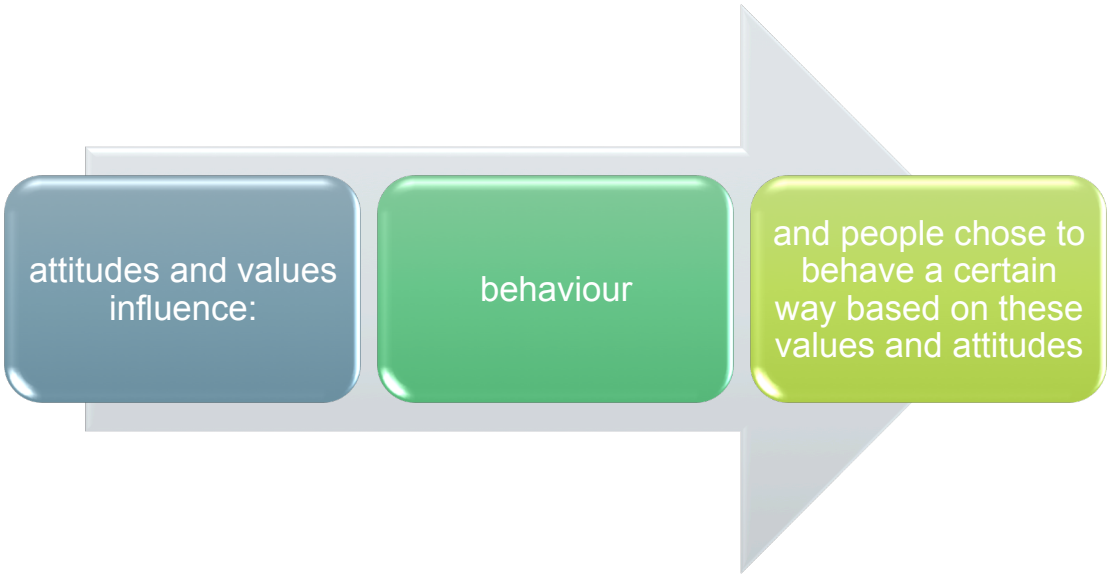


Figure 3: The ABC model based on Shove 2010

However, in practice people usually make more complex trade-offs between costs and gains (both financial and non- financial) and, consequently, the models are most likely not accurate. In Paul Stern's very influential ABC model (2002), C stands for context, not choice. And 'behaviour is an interactive product of personal-sphere attitudinal variables (A) and contextual factors (C)', - that is, more sociology-friendly factors such as regulations, technology, supportive policies. Stern made a very strong point that cognitive effort is key when people decide whether to engage in an efficiency action or not. Consequently, actual energy savings may differ a lot from modelled savings since misuse, fraud, rebound and neighbourhood effects have impacts on the real energy consumption (Mourik and Rotmann 2013). Estimated savings in evaluations are often higher than actual savings. When rebound (consumers increase the level of energy services after implementing efficiency measures) and pre-bound (energy consumption prior to the intervention is lower than the estimated energy consumption) effects are taken into account energy savings are often only half the estimated savings. Therefore, it is important to know how much energy is actually saved in order to make robust judgments of an energy saving intervention's effectiveness (Mourik and Rotmann 2013; Rosenow and Galvin 2013).

Critics furthermore, argue that sometimes, there is no reliable information available about the impacts on energy consumption for interventions (e.g. communication campaigns) and that it is therefore, not possible to calculate the cost-effectiveness. However, this does not necessarily mean that they have or have not been efficient (Breukers et al. 2009).

In addition, monitoring and evaluation of outputs and not outcomes will limit the ability to monitor and evaluate habitual behaviour. The House of Lords (2011) argues that the indicators and proxies currently used do not clearly reflect the real effects on habitual behaviours. Metered instead of modelled saving calculations are necessary to assess the real impact of the measures on energy consumption. Proxy indicators or indirect indicators[19] such as number of installed installations or kWh saved potentially are not even a real proxy; minor savings might involve most intensive behaviour changes whilst major savings might have been the result of a relatively isolated behaviour change, e.g. buying and installing a new heating system or LEDs[20].

A last point is that if only modelled savings are calculated, and real savings are not meeting these calculations, the uptake and acceptance of the involved technologies, e.g. passive houses, or services such as energy performance contracting will face serious problems (Batey, Mourik and Garcia 2013).

[19] We use the following definitions of proxies and indicators (retrieved from http://www.businessdictionary.com/definition/indicator.html#ixzz3Eo5JYZku): Proxy indicators are indirect measures or signs that approximates or represent a phenomenon in the absence of a direct measure or sign. Indicators are measurable variables used as a representation of an associated (but non-measured or non-measurable) factor or quantity.
[20] Comment by Ruth Rettie in concluding reflections at the 2014 Behave conference in Oxford

Figure 4 below, is ironically named 'a simplified diagram of a complex intervention' since it demonstrates the huge complexity of an intervention and the fact that behaviour change is not a linear process where behaviour A, if targeted with B, leads to C, leads to D and E. The diagram attempts to demonstrate that everything is interrelated and co-shaping and co-shaped Because of this complexity proxies such as savings or kWh reduction are unable to explain the why and how of behaviour change.

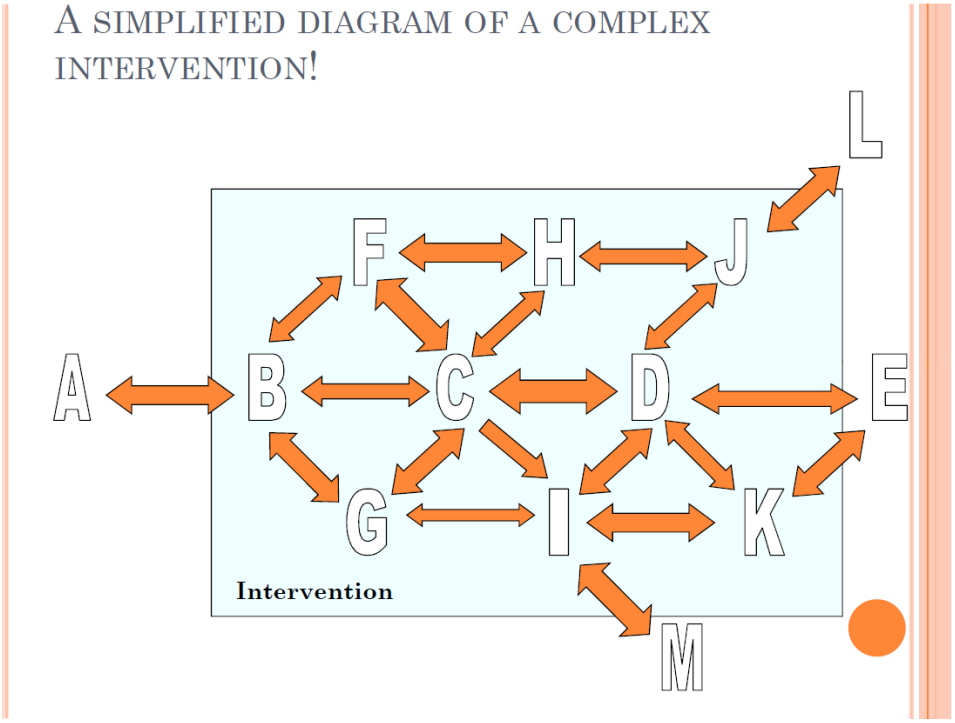


Figure 4: Pawson et al 2005.

Monitoring and evaluating distribution of costs and benefits: What are the costs and benefits to the end users?

As discussed above, monitoring and evaluation very often focus on calculating the cost effectiveness of an intervention. This calculation is usually based on dividing the actions performed, e.g. bought energy efficient installations, retrofitting etc. by the costs of the intervention as invested by the implementers. Or in other words, 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).[21]

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.

What is success according to whom: How do other stakeholders assess the successfulness of an intervention?

[21] See also IEA DSM Task 23 where the household soft costs of insulation were calculated.

As mentioned before, success of an intervention is usually evaluated on the basis of its cost effectiveness. However, this does not provide insights about whether or not long-term behavioural change is achieved and therefore, this indicator does not say much about the successfulness of an intervention in terms of realising lasting and actual behavioural change.

Cost-effectiveness also 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. Two different types of spill-over might be of particular interest, namely spill-over to:

i) other people, e.g., peers, neighbours, family and friends and
ii) other types of energy-related behaviour.

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. It could be valuable, in large national programmes such as insulation subsidy schemes, to do some pre-testing of what outcomes would mean a successful programme and to whom. For example, an Energy Minister would like to see maximum number of homes insulated for minimum tax-payer funded cost and have only good news stories in the media. The agency rolling out the subsidy scheme will need to measure and evaluate this, plus ensure that a healthy insulation market is created by creating training and auditing regimes and standards and labels (often working with other government departments and intermediaries, such as insulation installation providers). The insulation industry wants the subsidy to kick off a well-performing market, increase jobs and sell more insulation, often leading to a change in social or cultural norms where highly insulated homes become the standard (as in central and northern Europe but not in many other places, like New Zealand) and the industry can function on its own, without government subsidies. Home owners often want to hear about health benefits to their families, eg an improvement in respiratory illnesses leading to fewer sick days at school and saving money on bills in the long term which is something that may be of interest to trusted intermediaries, such as GPs or public health nurses who will collect information on eg reduced number of GP visits or incidence of respiratory illnesses.

There can be issues with a subsidy or loan scheme automatically leading to impressions of and conversations around financial gains, instead of concentrating on e.g. convenience, comfort or health improvements which may be of more value to many home owners. Also, some homes may experience a positive rebound effect by more effectively heating their homes to a higher degree, thus experiencing health and comfort benefits, but potentially slightly increased energy costs despite having insulation.

Of course a problem with focusing on multiple benefits for different stakeholders, not only intervention designers and funders also leads to the question of weighing the different (perceived) outcomes. In interventions that take a more comprehensive or systemic approach from the onset, with participation of multiple stakeholders, sometimes outnumbering energy stakeholders, the whole process of aligning all these interests and needs becomes a challenge in itself. A solid understanding of where the different stakeholders and actors in such a systemic intervention sit in terms of their perceptions of successful outcomes and the intervention meeting their needs, will help design interventions and their M&E regimes better from the outset. A collective impact approach, as studied by the Task 24 extension (Subtask 7), may go a long way to aid collecting and analysing these different mandates, drivers, needs and perceptions from the outset.

Monitoring individuals, not practices or collectives or multiple benefits: What are the effects of an intervention on collectives and systems?

The above discussion thus demonstrates that 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.

As discussed in earlier sections, the economic and psychological approaches dominant in the sustainable energy and environmental policy arenas mainly view consumption as an individual decision-making process. Sahakian and Wilhite (2014) argue that, consequently, interventions have focussed mainly on prompting individual behavioural change and propose as solutions more energy efficient technologies instead of focusing on collective behaviour or practices and developing the social context that is conducive to more efficient behaviour. However, as argued, since we increasingly appreciate the social context and the role of the collective in behavioural change, increasingly, the design of interventions starts to include elements of sociological theories (implicitly or explicitly).

One unit of analysis following a more sociological underpinning of interventions is a focus on lifestyle. Straightforward definitions of lifestyle are hard to find, but in this report we regard lifestyle as more than the conscious, intentional and routinised patterns of behaviour and consumption. Lifestyles can be regarded as 'social conversations' through we associate and/or differentiate ourselves from others. Since many of the signals in this conversation are mediated by goods, lifestyles are closely linked to consumption of goods and services and the material and resource flows in society. Hence lifestyles are connected to identity and status, which is partially expressed through the choice and display of things and behaviours, as symbolic codes (Edgar and Sedgwick 1999). Understanding these conversations, and understanding how behavioural change interventions may affect peoples' identity is helpful in achieving more effective interventions. It is particularly important to learn how to design interventions that target behaviour change but allow for people to maintain their identity and lifestyle, unless of course the lifestyle itself is based upon excessive consumption.

Monitoring and evaluating actual behaviour and the meaning it holds and qualitative insights from research with people themselves, is necessary to understand lifestyle and the standard practices that often are different in terms of duration and frequency than what people think when asked about it.

Another challenge related to the focus on the individual discussed above and focus on indicators of success only relevant to implementers and funders is that other outcomes such as 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.

Additionally, sometimes the narrow focus on outputs such as energy savings can lead to the labelling of an intervention as a failure, whilst other outcomes, sometimes at system level can actually lead to much higher energy savings, e.g. the decreased hospital visits, prevention of serious health conditions and associated increased productivity due to reduced sick days measured in the Warm Up NZ example. This sort of systemic understanding is seriously growing. Many authors state that, in order to be able to assess the effectiveness and efficiency of a DSM intervention the goals should be clear, not ambiguous and there should be consensus about the goals among stakeholders. However, in practice, intervention goals and instruments used may change during the design and implementation in reaction to larger contextual changes, and the changes in goals and instruments may be necessary to be able to continue the intervention. In addition, the design and implementation processes are often managed by multiple stakeholders and sometimes there is no consensus about the (changing) goals. Therefore, Breukers et al (2009) argue that in order to assess the successfulness of an intervention that has not met pre-set goals, it is important to also consider why these goals have not been met and whether this is necessarily an indication of failure.

Flexibility of an intervention can be conducive to success because it allows for changes to better match the intervention with the specific context or with changing or unforeseen circumstances. However, the flexibility of changing goals, aims and interrelatedness of issues makes it difficult to evaluate a highly flexible intervention with conventional monitoring and evaluation methods (Mourik

and Rotmann 2013). Therefore, the amount of flexibility an intervention demonstrates to react to context changes and issues can also be qualified as indicator for success and should be the focus of monitoring and evaluation of DSM interventions (Breukers at al. 2009).

No participatory process or feedback loops in the traditional M&E: End-users may also want to know the results of an intervention

A final challenge in current monitoring and evaluation practices is that these 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[22]. 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.

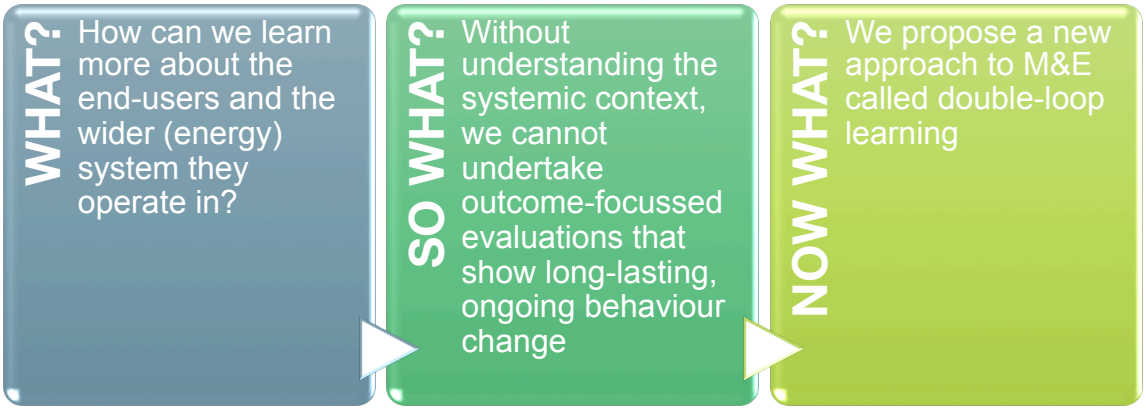
However, when designing an intervention that also accounts for end-user success indicators, that allows for an understanding of the for who, when, where, how, how long and why of behavioural change, and that accounts for the distribution of hard and soft costs and benefits; it is potentially very useful to involve end-users in both the design and implementation of the monitoring and evaluation. How end-users are represented or involved in the design process is of importance for the way in which an innovation is adopted (Breukers at al. 2009). In addition, 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). A crucial element in monitoring and evaluation is that people doing the monitoring and evaluation need to understand their own place and role in context.
These challenges show that monitoring and evaluating DSM interventions is no simple task. In the next section we will attempt to define what an alternative approach could look like.

No sharing = no learning = no mainstreaming: It is all about shared learning

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.
What we also need to realise is that most of the research and approach addressed in behaviour change is still very centered on residential area. It might be useful to investigate how monitoring and evaluating interventions that are designed for at home translate to the same people at work. There is this whole other side to behaviour when people are at work and much of our knowledge has not explored these differences and changed accordingly our monitoring and evaluation tools.
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.

[22] See Karlin et al's 2015 Methodological Review called Deliverable 3: How do we know what we know? IEA DSM Task 24.

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



So far we have attempted to show that 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 at 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. In addition, 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[23]. 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 as Shove stated[24].

[23] Reference Adam Cooper at the 2012 Task 24 Oxford Workshop
[24] Ppt of Shove at IEA DSM Task 24 Oxford 2012 conference

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.[25]

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 (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.

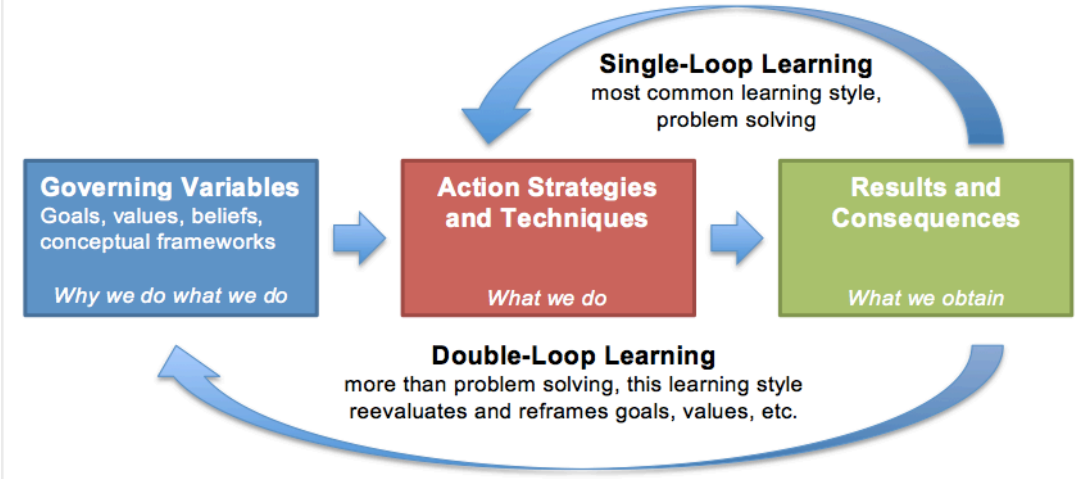


Figure 5: double vs single loop learning. Retrieved from http://www.afs.org/blog/icl/?p=2653

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 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.*

[25] Sarah Darby already argued for social learning in the context of the home in (Darby 2006)

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 at 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 the fact sheets in Deliverable 3B to provide more concrete guidelines for using single- and double-loop learning in M&E.

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

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

In the table below an overview of factors affecting behaviour change according to more socio-technical approaches are presented. These are all factors that should be monitored and evaluated in a more comprehensive or systemic intervention and for which relevant indicators or proxies should be identified as well.

Context 'factors'	How they affect opportunities towards lasting behavioural change

People	Practices and behaviours are affected by the people around us: direct peers like family, friends, neighbours, colleagues. In order to achieve long-lasting behavioural changes, it is important that peers also support or take up these new behaviours. Moreover, people learn best from other people so building social networks is important in DSM interventions. Stakeholders on a more distant level are important as well, e.g. policy actors who facilitate or inhibit change through policy support; or banks providing finance to new initiatives; energy companies etc. Intermediaries are also a very important group of people to incorporate in interventions.
Norms & Values, Culture	Practices are underpinned by norms which are socially-shared among smaller or larger groups of people. Changes in practices need to be supported by changes in social norms which provide the changed behaviours' legitimacy. Opportunities for change are affected by (local, regional, national) cultures, but cultures can of course also change due to changes in practices (over longer periods of time). Factors influencing cultural differences: learning culture; tradition and upbringing; risk attitude; prior experience of community engagement with similar interventions and/or developers; social cohesion/interpersonal relations; individual vs. group involvement; community trust; attitudes to new technology; privacy
Political factors	History of civic democratic engagement; types of government policies; stability of national policy; partisanship or collaborative governance (political culture); centralisation or federalisation of national government; tradition of top-down vs bottom-up initiatives; regulation and legislation.
Physical infrastructure	Urban and spatial infrastructure can inspire, encourage, constrain or even inhibit the uptake of more sustainable lifestyles. In cities, the uptake of healthier travel behaviour is not always supported by pedestrian-friendly or bike-friendly infrastructure. Physical infrastructure refers to all sorts of technologies, applications and products that are part of our daily lives and ways of doing (e.g. the short lifecycle of products limits possibilities to use these products sustainably).
Technology	Technology being implemented, the scale of a DSM intervention (large or small, centralised or decentralised, radical or incremental); technological flexibility and advancements; how technology fits into existing infrastructure
Geography	Options to behave more energy efficiently are constrained by climate, land availability, rural vs urban locations etc.
Socio-Economy	The overall economic situation affects peoples' daily lives, and ways of doing – and hence also opportunities for behavioural change (e.g. the need to save money may be a first trigger to change practices). Availability of natural resources and social acceptability of their exploitation; energy prices; technology and other input prices; perception of foreign investment; importance of energy independence; security of supply; interest in local employment and job creation; nationally-competing technologies and innovators.
Policy and implementation	Policy support is crucial and can either support or inhibit DSM interventions in several ways and on several levels. How is DSM implemented (community/local, regional or national level); organisational strength and make-up of policymakers and implementers; who are the right policy makers to talk to on each level; which agency makes the policy and which one implements it; how do they interact with each other

Table 4: Context 'factors' that affect behaviour change according to more socio-technical approaches positioning paper Oxford Task 24 conference, 2012).

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.[26]

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?[27]

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.[28]

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

[26] See Moezzi and Janda (2014) who criticised the current narratives in Energy Efficiency interventions and policy.
[27] Comments made by Janda at IEA Task 24 workshop in Oxford, September 5, 2014.
[28] The Task 24 extension will look into how this process can be both modeled and turned into real-life applicability by using a collective impact approach and a new model of the Energy System and the human players in it

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.[29] 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[30]. Additionally, the ICT

[29] See Batey, Mourik and Garcia (2014) for a more detailed discussion of this issue.
[30] Darby 2010, Hargreaves, Nye & Burgess 2010, Breukers & Mourik 2013

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: we need "metrics of resilience" that also reflect the need to keep our eye on the future as designers and evaluators.

This report, is as stated before, part two of three reports delivered under Subtask 3 of IEA DSM Task 24.

Epilogue

We are aware that this report 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 this 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, we appreciate that a key question for policymakers is in which way the different academic positions influence the design of an intervention. And how typical current interventions, assumed to be faulty because of a flawed theoretical understanding, should be changed or adjusted in order to better achieve the desired behavioural change. To answer these questions we will take a number of typical interventions from the country case studies, and "redesign" these interventions, including the monitoring and evaluation methods, according to the principles derived from the alternative theoretical understanding of the behaviour. This exercise will be conducted in Subtask 4 and be reported in the Subtask 4 country reports.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology 25, 273-291.

Annunziata, E., Rizzi, F., & Frey, M. (2014). Enhancing energy efficiency in public buildings: The role of local energy audit programmes. Energy Policy, 364-373.

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.

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., Mathijsen, T, Brunsting, S, Uyterlinde, 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.

Broennum, L., and Moller, S. (2013). The dynamics and facilitation of a living lab construct. Proceedings of The XXIV ISPIM Conference - Innovating in Global Markets: Challenges for Sustainable Growth Conference held in Helsinki, Finland on 16 to 19 June 2013. ISBN ISBN 978-952-265-421-2.

Brög, W., Erl, I., Ryle, J., & Wall, R. (2009). Evaluation of voluntary travel behaviour change: Experiences from three continents. Transport Policy, 281-292.

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.

Davies, N. (2012). What are the ingredients of successful travel behavioural change campaigns? Transport Policy, 19-29.

Edgar, A., Sedgwick P. (1999). Key concepts in Cultural Theory. Routledge.

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.

Fogg, BJ & Hreha, J (2010) Behavior Wizard: A Method for Matching Target Behaviors with Solutions. Proceedings of PERSUASIVE 2010, Copenhagen. http://captology.stanford.edu/wp-content/uploads/2010/10/Fogg-and-Hreha-BehaviorWizard.pdf

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.

Hazas, M., Bernheim Brush, A.J., & Scott, J. (2012). 'Sustainability does not begin with the individual.' Interactions 19, 5

Hoicka, C. E., Parker, P., & Andrey, J. (2014). Residential energy efficiency retrofits: How program design affects participation and outcomes. Energy Policy, 594-607.

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.

Joachain, H., & Klopfert, F. (2014). Smarter than metering? Coupling smart meters and complementary currencies to reinforce the motivation of households for energy savings. Ecological Economics.

Jones, B. D. (1999). Bounded Rationality. Annual Review of Political Science, 297-321.

Karlin, B., R. Ford et al (2015) How do we know what we know: Methodology Review for Assessing Behavior-Based Energy Interventions. IEA DSM Task 24 Deliverable 3.

Kobus, C.B.A., Mugge, R., Schoormans, J.P.L (2013). Wahing when the sun is shining! How users interact with a household energy management system. Ergonomics 56 (3), 451-462

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., & Rotmann, S. (2012). IEA DSM TASK XXIV: Closing the Loop - Behaviour Change in DSM: From theory to practice; Task Work Plan. Wellington, NZ; Eindhoven, NL: IEA DSM TASK XXIV.

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.

Murphy, L., Meijer, F., & Visscher, H. (2012). A qualitative evaluation of policy instruments used to improve energy performance of existing private dwellings in the Netherlands. Energy Policy, 459-468.

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 Builings, 450-458.

Raw GJ, Varnham J & Schooling J (2010) Focus on behaviour change - reducing energy demand in homes. CLG.
http://www.communities.gov.uk/publications/planningandbuilding/energybehaviourchange

Sahakian, M., & Wilhite, H. (2014). Making practice theory practicable: Towards more sustainable forms of consumption. Journal of Consumer Culture, 25-44.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive and reconstructive power of social norms. Psychological Science, 18(5), 429–434. doi:10.1111/j.1467-9280.2007.01917.x

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

Stern, P. (2002). New Environmental Theories: Toward a Coherent Theory of Environmentally Significant Behavior. Journal of Social Sciences. Volume 56, Issue 3, 407-424.

Strengers, Y. (2011) 'Designing Eco-Feedback Systems for Everyday Life'. Proceedings of CHI 2011. ACM, New York.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). Ecological rationality: Intelligence in the world. New York: Oxford University Press.

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.
Van Dam, S.S., Bakker, C.A., van Hal, J.D.M. (2012). Insighs into the design, use and implementation of home energy management systems. Journal of Design Research 10(1/2), 86-101.

Vassileva, I., Dahlquist, E., Wallin, F., & Campillo, J. (2013). Energy consumption feedback devices' impact evaluation on domestic energy use. Applied Energy, 314-320.

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

Yalçin-Riollet, M., Garabuau-Moussaoui, I., & Szuba, M. (2014). Energy autonomy in Le Mené: A French case of grassroots innovation. Energy Policy, 347-355.

IEA Demand Side Management Energy Technology Initiative

The Demand-Side Management (DSM) Energy Technology Initiative is one of more than 40 Co-operative Energy Technology Initiatives within the framework of the International Energy Agency (IEA).The Demand-Side Management (DSM) Energy Technology Initiative, which was initiated in 1993, deals with a variety of strategies to reduce energy demand. The following member countries and sponsors have been working to identify and promote opportunities for DSM:

Austria	Norway
Belgium	Spain
Finland	Sweden
India	Switzerland
Italy	United Kingdom
Republic of Korea	United States
Netherlands	ECI (sponsor)
New Zealand	RAP (sponsor)

Programme Vision: Demand side activities should be active elements and the first choice in all energy policy decisions designed to create more reliable and more sustainable energy systems
Programme Mission: Deliver to its stakeholders, materials that are readily applicable for them in crafting and implementing policies and measures. The Programme should also deliver technology and applications that either facilitate operations of energy systems or facilitate necessary market transformations

The DSM Energy Technology Initiative's work is organized into two clusters:
The load shape cluster, and
The load level cluster.

The 'load shape" cluster will include Tasks that seek to impact the shape of the load curve over very short (minutes-hours-day) to longer (days-week-season) time periods. Work within this cluster primarily increases the reliability of systems. The "load level" will include Tasks that seek to shift the load curve to lower demand levels or shift between loads from one energy system to another. Work within this cluster primarily targets the reduction of emissions.

A total of 24 projects or "Tasks" have been initiated since the beginning of the DSM Programme. The overall program is monitored by an Executive Committee consisting of representatives from each contracting party to the DSM Energy Technology Initiative. The leadership and management of the individual Tasks are the responsibility of Operating Agents. These Tasks and their respective

Operating Agents are:
Task 1 International Database on Demand-Side Management & Evaluation Guidebook on the Impact of DSM and EE for Kyoto's GHG Targets – *Completed*
Harry Vreuls, NOVEM, the Netherlands

Task 2 Communications Technologies for Demand-Side Management – *Completed*
Richard Formby, EA Technology, United Kingdom

Task 3 Cooperative Procurement of Innovative Technologies for Demand-Side Management – *Completed*
Hans Westling, Promandat AB, Sweden

Task 4 Development of Improved Methods for Integrating Demand-Side Management into Resource Planning – *Completed*
Grayson Heffner, EPRI, United States

Task 5 Techniques for Implementation of Demand-Side Management Technology in the Marketplace – *Completed*
Juan Comas, FECSA, Spain

Task 6 DSM and Energy Efficiency in Changing Electricity Business Environments – *Completed*
David Crossley, Energy Futures, Australia Pty. Ltd., Australia

Task 7 International Collaboration on Market Transformation – *Completed*
Verney Ryan, BRE, United Kingdom

Task 8 Demand-Side Bidding in a Competitive Electricity Market – *Completed*
Linda Hull, EA Technology Ltd, United Kingdom

Task 9 The Role of Municipalities in a Liberalised System – *Completed*
Martin Cahn, Energie Cites, France

Task 10 Performance Contracting – *Completed*
Hans Westling, Promandat AB, Sweden

Task 11 Time of Use Pricing and Energy Use for Demand Management Delivery- *Completed*
Richard Formby, EA Technology Ltd, United Kingdom

Task 12 Energy Standards
To be determined

Task 13 Demand Response Resources - *Completed*
Ross Malme, RETX, United States

Task 14 White Certificates – *Completed*
Antonio Capozza, CESI, Italy

Task 15 Network-Driven DSM - *Completed*
David Crossley, Energy Futures Australia Pty. Ltd, Australia

Task 16 Competitive Energy Services
Jan W. Bleyl, Graz Energy Agency, Austria / Seppo Silvonen/Pertti Koski, Motiva, Finland

Task 17 Integration of Demand Side Management, Distributed Generation, Renewable Energy Sources and Energy Storages
Seppo Kärkkäinen, Elektraflex Oy, Finland

Task 18 Demand Side Management and Climate Change - *Completed*
David Crossley, Energy Futures Australia Pty. Ltd, Australia

Task 19 Micro Demand Response and Energy Saving - *Completed*
Linda Hull, EA Technology Ltd, United Kingdom

Task 20 Branding of Energy Efficiency - *Completed*
Balawant Joshi, ABPS Infrastructure Private Limited, India

Task 21 Standardisation of Energy Savings Calculations - *Completed*
Harry Vreuls, SenterNovem, Netherlands

Task 22 Energy Efficiency Portfolio Standards - *Completed*
Balawant Joshi, ABPS Infrastructure Private Limited, India

Task 23 The Role of Customers in Delivering Effective Smart Grids - *Completed*
Linda Hull. EA Technology Ltd, United Kingdom

Task 24 Closing the loop - Behaviour Change in DSM: From theory to policies and practice
Sea Rotmann, SEA, New Zealand and Ruth Mourik DuneWorks, Netherlands

Task 25 Business Models for a more Effective Market Uptake of DSM Energy Services
Ruth Mourik, DuneWorks, The Netherlands

For additional Information contact the DSM Executive Secretary, Anne Bengtson, Liljeholmstorget 18,11761 Stockholm, Sweden. Phone: +46707818501. E-mail: anne.bengtson@telia.com
Also, visit the IEA DSM website: http://www.ieadsm.org