

Gained in translation: Evaluation approaches for behavioural energy efficiency programmes in the US and Canada

Dr. Sea Rotmann
CEO of Sustainable Energy Advice Ltd. and
Operating Agent of IEA DSM Task 24
43 Moa Point Rd.
NZ, 6022 Wellington New Zealand
Email: drsearotmann@gmail.com

Kira Ashby
Senior Program Manager, Behaviour
Consortium for Energy Efficiency
35 Village Rd., Suite 202
Middleton, MA 01949 USA
Email: kashby@cee1.org

Abstract

One of the greatest challenges of behavioural energy efficiency programmes in the US and Canada is verifying and getting credit for achieved energy savings. In 2018, the US Department of Energy (DOE), together with the Consortium for Energy Efficiency (CEE) and its 80 US and Canadian utility members, joined a global behaviour change collaboration through the International Energy Agency Demand Side Management Task 24 (IEA DSM Task 24) to tackle this difficult problem together.

This paper is co-authored by the Operating Agent for IEA DSM Task 24 and the National Expert for the US. This paper shares the learnings from project researchers and practitioners as well as an analysis of international best practice on evaluating behavioural programmes. The focus of the US work on this collaboration has been on better understanding which evaluation methodologies and approaches have strengthened the credibility of behavioural programmes in the US and Canada. This includes insights on various regulatory frameworks and barriers North American (US and Canadian) utilities are facing.

The project team also analysed studies on the persistence of energy savings after a behavioural programme has ended. In addition, the team began a preliminary assessment on the topic of engaging hard-to-reach (HTR) customers through behavioural efforts. We found that, even across North American utilities, there was a wide spectrum of approaches to what a 'behavioural savings programme' was and how it should be delivered and evaluated. One overarching insight was that a trusted relationship with the Regulator was key to ensure the implementation of a wide range of behaviour change programmes and pilots.

The work on this project took place in the US and Canada but also included insights from international Task 24 case studies. The objective of this effort was to synthesize international learnings to address the ongoing challenge of demonstrating for behavioural efforts what works, for how long, why, and for which energy users.

Introduction

The IEA DSM Task 24 aims to facilitate and share knowledge between multiple stakeholder sectors and develop recommendations about the influence of behaviour change in the effective implementation of energy efficiency policies and programmes. After a period of building a theoretical framework and collecting practical cases (Phase I),¹ Task 24 is now finishing the second phase (Phase II),² which has focused on engaging actual "Behaviour Changers" from five major sectors (decision-makers, experts, providers, middle actors, and the so-called conscience or third sector) in co-designing real life interventions. This included undertaking almost 60 country workshops with hundreds of Behaviour Changers in 17 countries; agreeing on a main topic of interest for participating countries; undertaking landscape and stakeholder analyses in these countries; supporting them with evidence-based scientific approaches and practical case study comparisons from other countries along the way; and designing behavioural interventions that were then implemented and evaluated in real life pilot projects.

¹ <http://www.ieadsm.org/task/task-24-phase-1/#section-8>

² <http://www.ieadsm.org/task/task-24-phase-2/#section-8>

Task 24 and US / CEE

The US, via the national expert CEE,³ has formally participated in Task 24 since the beginning of 2018. It joined Year 3 of Phase II of the Task. Since 2014, Project Partner See Change Institute⁴ (SCI), has provided in-kind support on the topic of evaluating behavioural interventions to the Task. Other participants in Phase II of Task 24 were Austria, Ireland, the Netherlands, New Zealand, and Sweden. The North American contribution (US and Canada) is funded by ten-member utility sponsors of CEE (eight US and two Canadian) and supported by two government agencies, US DOE and Natural Resources Canada (NRCan). The Task's national expert, Kira Ashby, heads the Behavioural Program at CEE.

The unique programme and regulatory context in North America helped shaped the focus area for Task 24 work in the U.S. and Canada. In these two countries alone, more than \$8 billion USD are spent annually on energy efficiency programmes, which far outpaces most other countries (CEE Annual Industry Report). These programmes are often, though not exclusively, run by investor owned utilities (IOUs), as opposed to Europe's government-run programmes. As a result, many IOUs are mandated to meet energy efficiency targets, and programmes that count towards these targets must meet cost-effectiveness and other evaluation requirements. Behavioural social science techniques can be leveraged to enhance savings from energy efficiency programmes. However, humans - and any energy usage changes caused by their behaviour - are much less predictable than energy-efficient equipment. Thus, estimating savings from behaviour is more challenging and therefore requires higher evaluation standards in the U.S. and Canada. As a result, the randomised control trial (RCT) was, and remains, the gold standard for evaluation of utility behavioural programmes. Yet not all programme approaches are evaluable via RCTs, and other rigorous and credible evaluation methods have not been widely accepted as alternatives. Consequently, many energy efficiency efforts that would be considered behavioural in other countries are not labelled as such in the U.S., and it can be difficult to demonstrate the value and gain approval for programmes that contain behavioural elements.

Despite these challenges, programme administrators in the U.S. and Canada continue to run over 100 programmes that include behavioural elements across the residential, commercial, and industrial sectors (CEE Behavior Program Summary). This paper concentrates on the US's chosen theme of "evaluating behavioural programmes: methodology, credibility, persistence and hard-to-reach customers." For an in-depth description of the second phase of Task 24, the overall methodological approach, and a detailed overview of the Subtasks, please refer to Rotmann and Ashby (2019).

For the purposes of the Task 24 project, a *Behaviour Changer* is a person or agency tasked with the goal of designing, implementing, evaluating and/or disseminating interventions geared at changing energy End-User behaviours. In Task 24, the team differentiates between five Behaviour Changer sectors: "the Decision-maker" (usually government on all levels), "the Provider" (usually energy- and energy technology-providing industry on all levels), "the Expert" (usually researchers and consultants from a multitude of disciplines, especially economics, psychology, sociology, and engineering), "the Conscience" (usually the third sector including NGOs, community organisations, consumer groups etc.) and "the Middle Actor" (usually service providers in direct contact with the end users).

Evaluating behaviour change interventions

The UK House of Lords (2011) stated that evaluation of behaviour change interventions is necessary to be able to assess four main 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 interventions at the population level.

Task 24 has long recognised (Mourik et. al., 2015) 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. This paper has 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.*"

The Task has assessed evaluation of behaviour change interventions, since 2012, in a variety of ways:

³ <https://www.cee1.org/>

⁴ <https://www.seechangeinstitute.com/>

1. General analysis of Evaluation, Monitoring and Verification (EM&V):
 - Mourik et. al. (2015a and b) provided an overview of how the three main behavioural disciplines (economics, psychology and sociology) approach EM&V, differently.
 - Karlin et. al. (2015a) undertook a methodology review of ten years of residential feedback literature to assess what instruments, if any, were provided in the publications for quantifying behavioural outcomes “beyond kWh.”
2. Case study analyses:
 - The Task 24 “Monster Report” (Mourik and Rotmann, 2013) analysed how over 40 case studies from 12 countries had used various models of understanding behaviour, and theories of change, in practice. Part of this analysis was a look at the various evaluation metrics used in the residential, transport, smart (feedback) technology, and small to medium business sectors.
 - Lang (2014) contrasted two EU research projects in the residential sector in Austria and how their different disciplinary approaches led to different outcomes.
 - Rotmann (2018a) evaluated the usefulness of a home energy audit tool (HEAT) kit in Auckland, New Zealand.
 - The Sustainable Energy Authority Ireland (SEAI, 2018) evaluated the usefulness of a similar home energy saving toolkit in Ireland.
 - Rotmann (2018b) also undertook a cross-country case study comparison of similar toolkits in ten other countries, provinces and states.
 - Cowan et. al. (2017 and 2018) used qualitative and quantitative data analyses to evaluate the impact of a building operator behaviour change programme in the largest health network in North America.
3. Specific development of evaluation tools:
 - Van Summeren et. al. (2015) provided templates for how to use a double-loop learning approach (see Argyris & Schön, 1978) in the residential sector on Energy Performance Certificates; mass marketing campaigns and subsidy programmes.
 - Batey and Garcia (2014), Batey et. al. (2014) and Batey and Mourik (2016) proposed how to undertake “do it yourself” evaluation using information and communication technology (ICT) tools.
 - Karlin et. al. (2015b and 2016) developed an evaluation tool based on the findings in their methodology review (2015a) that no standard method was used in the literature to evaluate behavioural interventions beyond kWh.
 - Southern California Edison (2016) further developed this “beyond kWh toolkit” using psychometric analysis.
 - Rotmann and Chapman (2018) successfully tested this toolkit using Bayesian modelling in the Irish Task 24 case study on home energy saving kits (SEAI, 2018).
 - Rotmann (2017) discussed how storytelling using a fairy tale story spin can be used to evaluate behavioural interventions.
 - Rotmann (2018c) provides templates for undertaking case studies, focus groups and interviews to support triangulation of quantitative and qualitative measures in a “Toolbox for Behaviour Changers.”
4. Regulatory frameworks and barriers to EM&V:
 - Kallsperger and Rotmann (2017) and the Grazer Energy Agency (2017) looked at how the current monitoring regime by the Austrian Regulator could claim up to seven times greater kWh savings from behavioural programmes if it was expanded to include non-energy benefits (NEBs).
 - Rotmann and Ashby (2019) analysed how the different regulatory regimes in North America (US and Canada) influenced behavioural savings programmes in the utility sector.

Even though Task 24 studied a wide variety of approaches to analysing, testing and developing evaluation strategies for behavioural interventions, for the purposes of this paper the focus will be on the North American (US and Canada) issue (Subtask 6) and Behaviour Changers (Subtask 7) described in more detail in Rotmann and Ashby (2019).

Methodology

A mixed-method model was used to analyse the chosen topic of “evaluating behavioural programmes: methodology, credibility, persistence and hard-to-reach customers” in the US and Canada. The various steps, deliverables and methods are outlined in Table 1 below. Subtasks 1-5 were part of Phase I of Task 24 and not part of this most recent study. The project timeline is shown in Figure 1.

Table 1. Main steps, deliverables and methods used in Task 24 in the US and Canada

Step	Deliverables	Methods
1	Identification of the top DSM issues in US (Subtask 6 “The Issues”)	Task 24 Workshop 1; internal CEE member discussion; survey of CEE sponsors
2	Identification of the <i>Behaviour Changers</i> in the US; national & international expert network (Subtask 7 “The People”)	Two Task 24 Workshops with CEE members and SCI; Task 24 expert network; BEHAVE and BECC Task 24 special sessions with international experts
3	Application and testing of Task 24 tools (Subtask 8 “The Tools”)	Two Task 24 Workshops in US; BEHAVE and BECC Task 24 sessions
4	Input for the final report (Subtask 10 “The US Report”)	Stakeholder and expert interviews and surveys; CEE Behavior Program Summary analysis; Task 24 case study analysis; input and feedback from workshop attendees; primary literature review

*Figure 1. Timeline of Task 24 / CEE collaboration.****Subtask 6 – The top issue in the US and Canada***

Task 24 participants honed in on the top issue for the US and Canada by several means. First, the team undertook a written survey of CEE member sponsors before the first workshop in April 2018. The survey included questions about current behavioural programmes, main barriers and opportunities, main topic interests, as well as expectations for this research collaboration. Participants further expanded on the issue definition during the first workshop (see Rotmann and Ashby, 2019) and chose “vampire (stand-by) energy waste in residential households” for our “Behaviour Changer Framework” exercise (see Rotmann, 2016 for description). After the workshop, the team provided a list of key issues which were highlighted as main topics of interest, with short summary descriptions, and the CEE sponsors voted for their preferred topic, selecting evaluation and credibility of behavioural programmes.

Subtask 7 – Behaviour Changers in the US and Canada

In Workshops 1 and 2, the team brought together the CEE sponsors, representing *Decision-makers* (DOE), *Providers* (the ten CEE utility sponsors), *Experts* (SCI, Task 24 and Rutgers Center for Green Buildings) and the *Middle Actor* (CEE). During the “Behaviour Changer Framework” exercise which is a visualisation tool to describe a given socio-ecosystem, including the end user behaviour, contexts and drivers; the various Behaviour Changers charged with changing the end user’s behaviour; and how well they all interrelate with one another (for full description see Rotmann, 2016), in Workshop 1 (see Rotmann and Ashby, 2019), it became clear that the team were missing the *Conscience*, such as the National Resources Defense Council (NRDC), also a CEE member. *Regulators* were also identified as missing decision-makers (see also Mintzberg, 1999) who determine which utility behavioural programmes can claim savings, how these savings should be evaluated and by whom (e.g. internal versus third-party evaluators). One former Regulator from California was later interviewed by the Task 24 Operating Agent to collect some of these important insights for the report.

In addition, the team relied on the large Task 24 expert network (Subtask 5), who attended and provided feedback at Task 24 workshops at two major international conferences – BEHAVE in Zurich and BECC 2018 in Washington, DC. Around 30 written feedback forms were collected from experts from almost 20 countries at

BEHAVE. Oral feedback was received at the BECC conference special session for Task 24. The team also conducted three written interviews with National Experts for Task 24 in Austria, New Zealand, and Sweden.

Results and discussion

Subtask 6 – The top issue in the US and Canada

Definition issues, regulatory frameworks and barriers – and how to overcome them

It became clear from the interviews with CEE sponsors (see Rotmann and Ashby, 2019) that definitions of what constitute behavioural savings programmes (and HTR customers) differ widely. Thus, their implementation and evaluation differed significantly across US states and Canadian provinces. These differences were largely due to disparate regulatory environments and whether the given programme administrator was a public entity or an investor-owned utility (IOU). Some utilities had very close and trusted relationships with their regulators and were able to perform and claim savings from a broader set of behaviour programmes and pilots. Others didn't necessarily have clear regulatory restrictions on specific behavioural programme types but still encountered challenges in obtaining regulatory acceptance for new and innovative behavioural pilots. Some, most notably Californian IOUs (see Draft White Paper, SCI 2017), were subject to stricter regulatory frameworks and requirements.

From the Californian Draft White Paper (SCI, 2017): “Initial definitions of behaviour-based programmes in California limited claimable savings to a narrow subset of behavioural interventions that employed: 1) Comparative energy usage, 2) Randomised control trials, and 3) Ex-post measurement. Thus, in practice, behavioural savings programmes in California have been primarily focused on Home Energy Reports (HERs) that employ comparative feedback (providing energy end users with information about how their usage compares to that of similar households) – this is particularly true for investor-owned, as opposed to publicly-owned, utilities. However, utilising a broader set of social science techniques in more innovative behavioural programmes could help tap into the estimated 20 percent savings potential from residential behavioural energy efficiency programmes.”

Most sponsors said that it all depends on how behaviour is defined, but on average, organisations offer one to two behavioural programmes and usually they can claim savings for all their programmes. The most-noted programme was HERs, with six organisations running HERs (although two of them were discontinued due to low energy savings), followed by three Strategic Energy Management programmes. Other programmes reported included three programmes claiming energy efficiency widgets sent to their customers, two Apps, and one major customer loyalty programme. Additional programmes, such as school programmes or energy saving kits were mentioned by sponsors, but their savings (in at least one case, quite substantial: one school in a competition reduced their energy usage by 18 percent) typically could not be claimed.

One of the CEE member organisation participants clarified the challenges with behavioural programme definitions: “*It all depends on how you define “behaviour.” Energy efficiency or other parts of the utility that look at energy saving kits or school programs are often not included, same with education and training programs. You would say everything we do is behaviour, and we’d say hardly anything is – by our Regulator’s definition.*”

Contrast this quote, and the narrow scope of a behavioural programme as defined in California (see above), with the very broad definition of Task 24,⁵ which would certainly encompass all the behavioural, energy efficiency, and DSM programmes mentioned by the CEE sponsors.

It also became clear that as much as some IOUs operated under specific regulatory requirements, other IOUs had very few, if any, issues related to regulatory oversight. Broader behavioural programme acceptance issues that were mentioned by the sponsors, and which are supported by the analysis of Sovacool, 2014, included:

- Lack of basic programme understanding by both regulators and within the utility (two of the ten interviewees mentioned this)
- Stakeholder perceptions of behavioural techniques (two times)
- Acceptance of new methods and qualitative analyses for proving cost-effectiveness (four times)
- Budget constraints (two times)

⁵ *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, word-of-mouth 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-side management but should be identified and specified by the Behaviour Changer designing the intervention for the purpose of outcome evaluation.

In terms of possible solutions to some of the barriers mentioned above, CEE sponsors (four) recommended to start small with pilots to show proof of concept and then scale up. Other responses included:

- Educational meetings and build trusted relationships with the Regulator (five of the ten interviewees mentioned this)
- Looking to other countries for insights (two times)
- Including NEBs or Non-Energy Impacts (NEIs) (two times)

CEE Behavior Program Summary

The annual CEE Behavior Program Summary⁶ (BPS) compiles information on CEE member programmes from throughout the US and Canada. It serves as a way to share information on programme details, behavioural insights incorporated into programmes, evaluation design, savings, and learnings. Since the annual data collection began in 2012, 279 programmes have been reported from 94 organisations representing 41 US states and three Canadian provinces. A majority of the programmes reported have been from the residential sector (76 percent), with the commercial and industrial sectors representing 24 percent and 11 percent, respectively. Given that the data used in development of the BPS is self-reported and is neither a census nor a representative sample, it is important to note that the programmes it includes may not be representative of all CEE member behavioural programmes in the US and Canada.

The BPS captures which programmes include the following behavioural insights (see Figure 2). Of all the social science techniques listed, programmes reported using the following most frequently: Feedback (providing energy usage or related cost information to end users), Social Norms (demonstrating to customers how their energy usage compares to that of similar individuals), Modelling Behaviour (demonstrating or illustrating the desired energy usage behaviours) and Prompts (reminders to take specific actions). Using trusted Messengers, Interpersonal Communication, Priming and Goal Setting are also relatively common strategies.

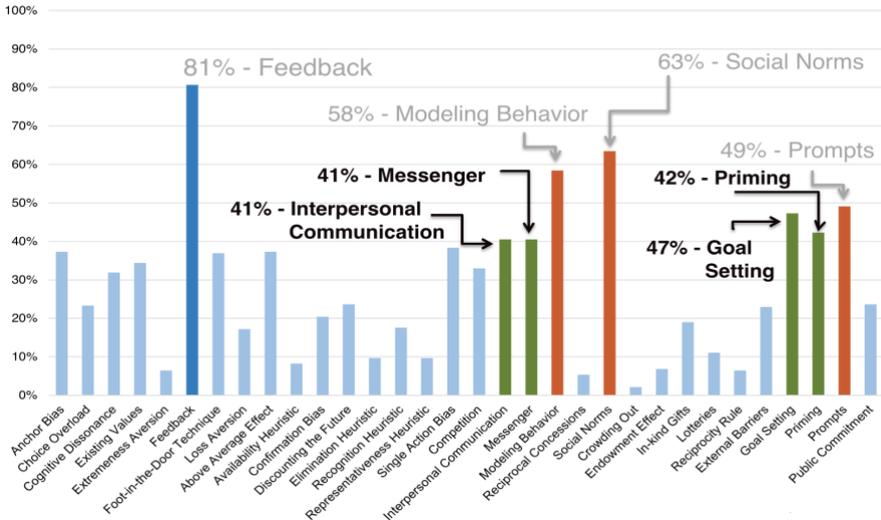


Figure 2. Most reported behavioural insights, CEE BPS. The colours indicate the relative tier of frequency with which different behavioural techniques were reported being used in programmes (e.g. most commonly reported behavioural techniques are shown in darker blue, whereas the least common are in lighter blue).

Evaluation is an important aspect of programme implementation and can significantly impact whether or not a utility can claim savings - in other words, count any achieved energy savings from the programme toward their utility’s total required energy efficiency targets - and use behaviour-based approaches. Thus, the BPS also collected data on evaluation. Of the 279 programmes reported, 44 percent had completed a programme evaluation at the time of reporting, while 22 percent had an evaluation planned or still underway. For the programmes that had either completed evaluations or had evaluations planned, a majority of those programmes used impact evaluations (97 percent of programmes) with just over half receiving process evaluations (60 percent of programmes; see Figure 3). The evaluation metrics collected in the BPS included: energy savings, customer satisfaction, awareness, bill savings, and number of participants reached.

⁶ <https://library.cee1.org/content/2018-behavior-program-summary-public-version>

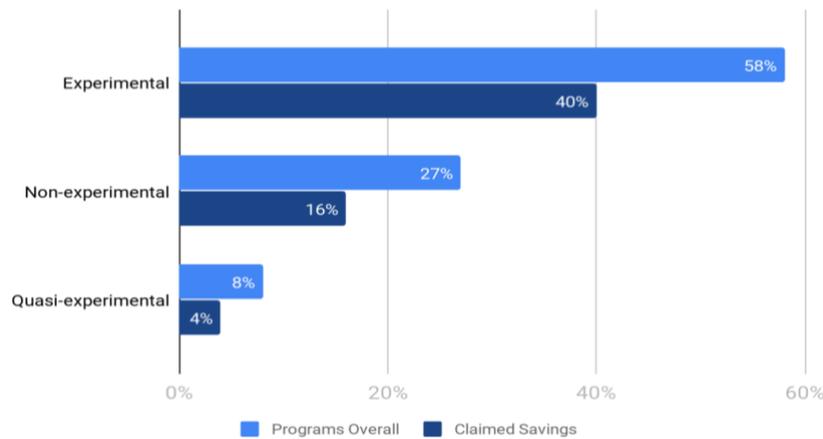


Figure 3. Types of evaluations for CEE member programmes reported in BPS. For the purposes of this data collection, an experimental design includes random assignment of participants to the treatment and control groups; quasi-experimental design includes a treatment and control group but no random assignment; and a non-experimental design includes neither random assignment nor a treatment and control group distinction.

Credibility of evaluation methods

It became clear from the sponsor discussions around what constitutes a behavioural programme, that most utilities did indeed undertake many other interventions, which under the Task 24 definition would be called “behavioural.” However, these other programmes are often undertaken by different teams within the utility and may have different evaluation methodologies from those programmes the utility considers to be behavioural. Third-party evaluators are commonly used to undertake all evaluations, though deemed measures are less common and often depend on white papers published by the regional energy commissions.

A majority of the ten CEE members interviewed (six) use randomised control trials (RCTs) for claiming savings; one member uses deemed savings and two use customer modelling. Another two mentioned randomised encouragement designs as alternatives to randomised control trials (see West et al, 2008). Several interviewees expressed interest in using behavioural and qualitative insights more extensively in order to undertake and scale up more ambitious pilots.

During an interview for this project (see Rotmann and Ashby, 2019), the former Californian Regulator stated the importance of “not picking evaluation methodology winners.” She has articulated her experiences from eight years at the California Public Utilities Commission (CPUC) in a lengthy blog,⁷ which also addresses the newer Pay for Performance and Normalised Meter Energy Consumption (NMEC) programmes and cautions all players to learn from history. She writes that: *“All actors in the system need to know how they will be measured and judged to assume responsibility and maintain accountability, and they need the ability to monitor progress as things roll out. Methods need to be consistent, but also transparent and replicable. To be replicable, the nitty-gritty execution of the method must be accessible. If there is too much room for interpretation, methods will be vulnerable to persistent attack, usually from the party that didn’t get a favourable outcome. Energy efficiency requires upfront agreement, because there is no other ground truth. Allowing energy efficiency savings calculations to be endlessly disputed make scaling the system difficult, impedes efforts to create and exchange a commonly understood unit of savings, and sacrifices the general sanity of all who are involved. Citing a generic method is not enough. Publishing and following evaluation frameworks and protocols is not enough. The method and the execution of that method must be accessible at the design phase, foundational to deployment of the effort, and serve as the contractual basis for payment. There can be no shortcuts here. Having qualified evaluators on either ‘side’ is not enough, as that only ensures the debate will be robust and last forever.”*

Persistence and HTR customers

Most sponsors (six) said they had not yet measured persistence. Persistence is the rate at which behaviour changes and related energy usage reductions achieved by a programme shift back towards previous baselines following the conclusion of a programme. Many sponsors indicated an interest in better understanding persistence, but hadn’t typically studied it because their behavioural programmes were too new. Three organisations noted that they were measuring persistence and have shared their reports with CEE. CEE (Ashby et. al., 2017) published a synthesis of members’ findings on persistence. A review of the primary literature showed that out of 39 papers deemed directly useful to the CEE sponsors, 19 (or almost half) measured persistence in some way.

⁷<http://www.zondits.com/article/15763/the-promise-of-performance>

On the topic of programmes aimed at “hard to reach” (HTR) audiences, most sponsors said that their behavioural programmes do not typically target HTR customers specifically, although this is mainly due to definitions of what constitutes a “behavioural programme” and a HTR customer (see above). Several expressed interest in this topic for future pilots. Their varying definitions and observations about HTR customers included:

- If you can send a bill or deliver electricity to a customer, they are not HTR, by definition.
- It is all about getting the right people involved; finding out who the decision-makers are can be hard.
- They are more “hard to serve,” such as really low income, renters, and multifamily apartment buildings.
- Rural, non-white and low to moderate income customers were regarded as HTR by one sponsor.
- Customers who don’t participate in behavioural savings programmes (ex: renters or landlords).
- Some simply called them “underserved.”

According to the CPUC Regulator (2013), any two of these criteria are sufficient to constitute HTR: they don’t have easy access to programme information, they don’t participate in energy efficiency programmes, there is a language barrier, business size barrier, geographic barrier, lease barrier, etc.

The finding that HTR customer definitions vary widely even within the utility sector and one geographic region (the US and Canada) led to the development of a new IEA DSM Task to follow Task 24 that focuses on the HTR energy user groups in the residential and commercial sectors (see Rotmann, 2019).

Subtask 7 – Behaviour Changers in the US and Canada

The ten CEE sponsors and collaborators represent all major US regions and two Canadian provinces, in addition to US and Canadian decision-makers (DOE and NRCan). Given the key role the sponsors played in the project, their representation of 24.6 million electric customers and 8.5 million natural gas customers throughout the US and Canada was essential to this paper’s findings. The most interesting findings in the final report (Rotmann and Ashby, 2019; see discussion below) came from the interviews with CEE sponsors and other experts, including the former CPUC Regulator.

International comparisons

From the literature review, international expert interviews, and Task 24 workshops and special sessions at the BECC and BEHAVE 2018 conferences, the research team made the following findings:

- There is a lot of primary literature on evaluation, but as our Methodology Review by Karlin et. al. (2015a) has shown, not much of it clearly outlines evaluation methodologies, nor do many papers give the actual survey questions or scales that were used for qualitative analysis.
- In addition, primary literature is often not relevant to those running energy efficiency programmes at utilities, as the work undertaken is usually more academic and smaller in scale than most US and Canadian utility programmes that often target tens of thousands of customers.
- Studies on HTR energy users are harder to find in the primary literature than more general behaviour change intervention or persistence studies.
- In terms of CO₂ emissions, behaviour of individuals in the US has been found to be the single-highest emissions sector (Vandenbergh et. al., 2010), yet has received much lower research funding relative to other energy efficiency topics such as energy-efficient technologies. This is also reflected in the amount spent by utilities on behavioural savings programmes versus broader energy efficiency or demand response programmes.
- Even though the EU and US both regulate the utility sector to implement energy efficiency and behaviour change interventions and demonstrate energy savings via prescribed methodologies, there are some specific differences in how this is done, particularly around consequences of non-compliance (in the US, it is more incentive-based whereas in the EU it tends to be more penalty-based) and motives (in the US, it seems more related to financial incentives, in the EU more to moral considerations). A country like New Zealand, with a highly-deregulated utility sector, has no comparative regulations, only government-set strategies and targets, with little consequences for non-compliance. New Zealand utilities thus have less motivation to extensively evaluate any behavioural interventions or pilots.
- Given the large variability even across US states in terms of regulatory frameworks and what constitutes a behavioural programme, any cross-country comparison has to be taken with a grain of salt. In looking from the US and Canada to other countries, it is important to consider and understand the different underlying contextual factors.
- There are significant differences in research approaches to behaviour change in the EU and New Zealand as opposed to the US and Canada, particularly around the emphasis on individual versus societal behavioural changes: This includes different disciplinary approaches and models, e.g. a preponderance of psychological theories such as the *Theory of Planned Behaviour* (Ajzen, 2011) in the US versus sociological theories such as *Practice Theory* (e.g. Reckwitz, 2002) in the EU and also more

social-marketing focused approaches like *Community-Based-Social-Marketing* (CBSM, McKenzie-Mohr & Schultz, 2014) or neoclassical economic models like the *Deficit Model* (Simis et. al., 2016) in the US. However, in both the EU and US, behavioural economics and behavioural insights have found a strong niche and support by policymakers and industry alike (despite its relatively limited application in the energy sector, e.g. Mourik & Rotmann, 2013).

- Even though European experts seem to have a stronger mandate to undertake “purely” behaviour change research and pilots, they still face some of the same challenges around credibility that the CEE sponsors have raised. These challenges include lack of funding and support for behavioural programmes as compared with energy-efficiency technology research and technology programmes. Other challenges are around engineers and policymakers being dubious about the validity of qualitative data and evaluation. At the 2018 BEHAVE conference Task 24 workshop, the team also had an interesting discussion around social justice and class issues, particularly when addressing the “HTR” or fuel-poor sector (see Rotmann & Ashby, 2019).
- A lot of these issues can be overcome by following proper social science and design thinking processes, which can be internationally-validated and become something of a “standard” of how to undertake and evaluate behaviour change interventions. SCI workshoped such a process with our CEE sponsors, using Task 24 case studies to illustrate both best practice – and less effective practice – in real-life interventions (see Rotmann and Ashby, 2019).

Conclusions

The following main regulatory issues and insights emerge for the US and Canadian utility sector, as evidenced by observations collected through both the CEE BPS and the CEE sponsor interviews:

- There are different definitions of what constitutes a “behavioural savings programme” – the Task 24 definition is most broad and the CPUC 2013 definition is the most limited. Most CEE sponsors’ programmes lie on a spectrum in between. Although “behaviour is everything” is arguably true, it is also rather lofty and thus difficult to act on in the “real world”.
- Most sponsors undertake other energy savings programmes that do have behavioural aspects - such as programmes where widgets are deemed or demand response programmes using time-of-use tariffs. Several have extensive school and energy literacy programmes, the savings of which cannot currently be claimed in some jurisdictions.
- How to measure and evaluate behavioural savings programmes is not always straightforward. In some states, RCTs are the primary or only acceptable evaluation methodology, while in others programme administrators only need to prove “cost-effectiveness,” and the methods allowable in these cases are typically more flexible.
- The use of third-party evaluators and third-party implementers is common, particularly in California where it is increasingly required by regulators, although in-house EM&V expertise is regarded as cost-effective and desirable by many utilities.
- A close, trusted relationship with the Regulator is also highly desirable and can lead to much greater flexibility in terms of what can be piloted or claimed as behavioural interventions. Perhaps co-creation of new programmes in which both programme administrators and regulators collaborate on programme development could help facilitate buy-in from both parties upfront. Trust needs to be (re)built to enable collaboration between different actors in the system, including the utilities and their customers, regulators, implementers and evaluators.
- Intra-organisational barriers are also of concern, especially around demonstrating that qualitative or behavioural measures are of value and can be triangulated to improve insights from quantitative data; time, budget, and resource constraints can also be a challenge; and “re-inventing the wheel” rather than relying on what has been shown to be working elsewhere. Small-scale pilots are invaluable here.
- Proving persistence of energy savings is something of a holy grail – many sponsors are interested in it, but most have not yet had the chance to measure the persistence from their programmes. The recent 2017 CEE report on persistence is still the main standard, but further research is necessary.
- HTR energy users encompass many sub segments, which need to be engaged differently. There is a huge variety of different HTR definitions, even within utilities in the U.S. and Canada. HTR definitions and segments also vary widely across the residential and commercial sectors. The term “underserved” may be a more holistic description of the many user groups that fall into the broader HTR category, if current behavioural or energy efficiency programmes fail to engage them. The IEA DSM Programme is in the process of developing a new Task, following Task 24, that focuses on the “HTR Energy Users” in the residential and commercial sectors. The team welcomes any interested experts or countries who want to partake in this new international research collaboration on behaviour.

Acknowledgements

With many thanks to Project Partner Dr. Beth Karlin from the See Change Institute, CEE staff Veronica Alix and Jennifer Smith, and our 11 CEE member sponsors who made U.S. participation in Task 24 possible.

References

- Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology and Health* 26 (9), pp. 1113-27.
- Argyris, C. and Schön, D. (1978). *Organizational learning: A theory of action perspective*. Reading, Mass: Addison Wesley.
- Ashby, K. CEE Behaviour Program Summary (2012-2018).
- Ashby, K., Gutierrez, V., Menges, S. and J. Perich-Anderson (2017). Keep the Change: Behavioral Persistence in Energy Efficiency Programs. IEPEC conference, Baltimore, MD.
- Batey, M., and Garcia, J.P. (2014). Innovative Use of ICT to uncover space-usage patterns & improve energy-efficiency. EESAP5 Proceedings.
- Batey, M., Mourik, R.M., and Garcia, J.P. (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. IEPEC conference, Berlin, Germany.
- Batey, M., and Mourik, R.M. (2016). 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, *Energy Efficiency* 9(4), pp. 939–950.
- Cowan, K., Sussman, R., Rotmann, S. and Cox, R. (2017). CHS case study: Designing a successful behaviour change programme for hospital building staff. IEA DSM Task 24.
- Cowan, K., Sussman, R., Rotmann, S. and Mazzi, E. (2018). It's Not my Job: Changing Behavior and Culture in a Healthcare Setting to Save Energy. ACEEE Summer Study Monterey, US.
- Californian Public Utilities Commission (2013). *Energy Efficiency Policy Manual, Version 5*. CPUC R.09-11-014, pp. 64.
- Grazer Energieagentur (2017). *Ergebnisbericht [in German]*. IEA DSM Task 24.
- Kallsperger, T. and Rotmann, S. (2017). Subtask 6&7 - Final Report Austria. IEA DSM Task 24.
- House of Lords (2011). *Behaviour Change*. Retrieved from: <http://www.parliament.uk/business/committees/committees-a-z/lords-select/science-andtechnology-committee/inquiries/parliament-2010/behaviour/>
- Karlin, B. et. al. (2015a). Subtask 3 Deliverable 3. IEA DSM Task 24.
- Karlin, B., Ford, R. and McPhearson-Frantz, C. (2015b), *Exploring Deep Savings: A Toolkit for Assessing Behavior-Based Energy Interventions*, IEPEC Conference, Long Beach, US.
- Karlin, B., Ford, R., and McPhearson-Frantz, C. (2016), *Evaluating Energy Culture: Identifying and validating measures for behaviour-based energy interventions*, IEPEC Conference, Amsterdam, Netherlands.
- Lang, G. (2014). Subtask 2 - The Energy Hunt Austria. IEA DSM Task 24.
- McKenzie-Mohr, D. and Schultz, W. (2014). *Choosing Effective Behavior Change Tools*. *Social Marketing Quarterly* 20(1).
- Mintzberg, H. (1999). „Who should control the operation? In: Mintzberg, H., Quinn, J.B. & Ghoshal, S. „The Strategy Process“. Harlow: Pearson Education Ltd
- Mourik, R.M. and Rotmann, S. (2013). Subtask 1 - Most of the time what we do is what we do most of the time. And sometimes we do something new. IEA DSM Task 24.
- Mourik, R.M., van Summeren, L., Breukers, S. and Rotmann, S. (2015a). Subtask 3 Deliverable 3A - Did you behave as we designed you to? Subtask 3, Deliverable 3A: Positioning paper on monitoring and evaluation. IEA DSM Task 24.
- Mourik, R.M., van Summeren, L., Rotmann S., and Breukers, S. (2015b). Did you behave as we designed you to? Monitoring and evaluating behavioural change in demand-side management: from what to why. ECEEE Summer Study 2015.
- Reckwitz, A. (2002). Toward a Theory of Social Practices: A Development in Culturalist Theorizing. *European Journal of Social Theory* 5 (2).

Rotmann, S. (2016). How to Create a ‘Magic Carpet’ for Behaviour Changers, BEHAVE Conference, Coimbra, Portugal.

Rotmann, S. (2017), “Once upon a time...” Eliciting energy and behaviour change stories using a fairy tale story spine, Energy Research and Social Science, Special Issue 31 on Narratives and Storytelling in Energy and Climate Change Research.

Rotmann, S. (2018a). Subtasks 6&7 - Evaluating Auckland Council's HEAT kit. IEA DSM Task 24.

Rotmann, S. (2018b). Subtasks 6&7 - Cross-country case study comparison of home energy saving kits. IEA DSM Task 24.

Rotmann, S. (2018c). Subtask 8 – Toolbox for Behaviour Changers. IEA DSM Task 24.

Rotmann, S. and Chapman, D. (2018). Subtask 9 - Evaluation Report for Home Energy Saving Kits: Using Bayesian Modelling to test the “beyond kWh” toolkit in Ireland. IEA DSM Task 24.

Rotmann, S. and Ashby, K. (2019). Subtasks 6&7 - Final US Report: IEA DSM Task 24.

Rotmann, S. (2019). Task Definition “Hard to Reach Energy Users”. IEA DSM.

See Change Institute (SCI, 2017). Draft White Paper to the Californian Public Utilities Commission.

Simis, M.J., Madden, H., Cacciatore M.A. and Yeo, S.K. (2016). The lure of rationality: Why does the deficit model persist in science communication? Public Understanding of Science 25(4).

Southern California Edison (2016). Subtask 9 - Dimensions of Energy Behavior: Psychometric Testing of Scales for Evaluating Behavioral Interventions in Demand Side Management Programs. IEA DSM Task 24.

Sovacool, B.K. (2014). What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. Energy Research & Social Science, 1. pp. 1-29. ISSN 2214-6296

Sovacool, B. K. and Hess, D.J. (2017). Ordering theories: Typologies and conceptual frameworks for sociotechnical change. Social Studies of Science 47(5).

Sustainable Energy Authority Ireland (SEAI), (2018). Subtask 6&7 - Final Report Ireland. IEA DSM Task 24.

Vandenbergh, M.P., Stern, P.C., Gardner, G.C., Dietz, T. and Gilligan, J.M. (2010). Implementing the behavioral wedge: designing and adopting effective carbon emissions reduction programs. Environmental Law Reporter, 40, pp. 10547-10554.

West, S.G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D.G., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M., and P.D. Mullen, (2008). Alternatives to the Randomized Controlled Trial. Am J Public Health 98(8): 1359-1366.