ENERGY EVALUATION EUROPE
2020 Conference
Here East, Olympic Park, London
Monday 29 June to Wednesday 1 July 2020
Formerly IEPPEC
Accelerating the energy transition for all: Evaluation's role in effective policy making

How to learn fast? A case study of challenges in evaluating a programme designed to bring about systems transformation

Michelle Hollier, Winning Moves, UK (michelleh@winningmoves.com)
Emma Harrison, Energy Systems Catapult, UK

ABSTRACT

Through the Smart Systems and Heat (SSH) Programme, the Energy Systems Catapult (ESC) delivered a large consumer-focused project aimed at identifying how to overcome the barriers to the decarbonisation of residential heat. SSH was designed as an opportunity for learning about what and how to innovate in the delivery of residential heat.

Evaluation was built into the programme, with a role in helping to deliver this learning aim. This paper will use the evaluation of SSH as a case study of challenges in evaluating what has and will work with regards to systems innovation and the methods and approaches that do, don't and might be used to address this. In particular:

- How to identify relevant outputs and outcomes? The understanding of what the programme could do (within the context of the wider decarbonisation agenda) developed throughout delivery.
- How to engage with programme stakeholders in an appropriate and timely way? Stakeholder groups adapted through the programme and the messaging of the programme evolved during delivery through testing and refinement with stakeholders.
- How to provide useful learning about the potential of approaches to systems innovation? The evaluation was conducted concurrent to delivery of SSH2; this limited the extent to which the achievement of intended outcomes could be accurately assessed.

We need to learn fast about what can work (and won't work) to understand how to achieve a transition in the energy market. This paper reflects on the challenges for the evaluation practitioner in supporting this.

Introduction

Heating accounts for 37% of total UK carbon emissions, with heating for our homes and buildings responsible for around 20%. To achieve our Net Zero targets, the UK's 27 million households will need to rapidly adopt new low carbon heat solutions through the 2020s and 2030s.

The Smart Systems and Heat (SSH) programme was designed to help innovators address this market gap and unlock the commercial opportunity of low carbon heating, by:

1. Identifying how to address the technical, regulatory, economic and social barriers that block new low carbon heat products, services and business models getting to market;
2. Establishing a range of platforms, insights and modelling tools to help innovators discover new low carbon heating solutions that consumers value and;

3. Bringing innovators, businesses, local authorities, networks, policy-makers, regulators and consumers together to understand how to create new markets that deliver low carbon heating solutions at scale.

In these efforts, the programme was trying to understand what it would take to achieve systems innovation; innovation that aims to change the underlying structure of a system rather than any of its individual parts. Systems innovation is a term that has risen to prominence recently as societies, governments, enterprises, and professionals find themselves facing a new set of highly complex challenges (also referred to as "wicked problems"). Climate change is one of these 'wicked problems'. Recognising the interconnected and complex nature of the energy system, the programme sought to inform our understanding of how we can build a better energy system to meet 2050 targets.

The Energy Technologies Institute (ETI) launched SSH in 2012 and funded Phase 1 of the programme, which was delivered by Energy Systems Catapult and its partners. The substantive elements of Phase 1 of the programme were delivered by 2016, with some outputs delivered subsequent to this.

Smart Systems and Heat Phase 2 (SSH2) was a successor programme to Phase 1 and took the form of a £9.8m grant from the Department of Business, Energy and Industrial Strategy (BEIS) to fund activity between Summer 2017 and March 2019. The programme was set up to develop evidence on how to facilitate the take up of low-carbon heating technologies across the UK through the development of new digitally enabled services.

The aim of SSH Phase 2 was to deliver a test bed with local authorities and industrial partners (delivery partners) for understanding the needs and desires of discerning energy consumers in order to develop and evaluate examples of potential new service offerings, as well as developing the understanding of the related policy implications.

There were three work packages within the SSH2 programme:

A winter trial of households equipped with advanced heating controls. A 'living lab' of 108 homes was created to understand consumer energy use and preferences and test the concept of 'Heat as a Service' during winter 2017/18. As part of the trial, an advanced heating control system was installed in participants' homes. The system provided room by room control of the heating. Participants used the equipment to control heating in their homes during the winter while data was collected about their heating requests and usage patterns. After participants had become accustomed to using their new systems, they were introduced to the concept of a heat plan. A heat plan allowed consumers to buy a 'warm hour' (heating used for an hour), rather than the traditional method of paying by unit of fuel used. Participants were then offered the choice to transfer to a heat plan, allowing investigations into how people would react to this opportunity. These heat services were an initial attempt to mimic the services that may be available in a future energy service market. Throughout the winter trial, participants were asked to complete various research activities in order to provide an understanding of participants' experiences of HESG and the 'Heat as a Service' offer. A wide range of data were generated from these research activities and the HESG platform itself.

Business modelling. A series of market engagements, research and analyses were conducted to explore consumer-focused delivery of heat decarbonisation. The outputs of the work included:

- A back-casting exercise - analysis which starts from a desired future state (in this case a new market) and 'back-casts' to identify how a journey to get there might unfold. It was used to layout a potential market transition pathway
- Industry workshops - with incumbents, new entrants, and consumer advocates to gauge opinion and elicit feedback on the key issues and potential solutions generated in the back-casting activity to identify a potential pathway to market transformation
- Consumer research - surveys to understand initial reactions to a range of potential policy interventions and exploring consumers' preferences when placed in the scenario of navigating a transition to low carbon heating.

Local engagement. ESC worked collaboratively with Greater Manchester Combined Authority, Bridgend County Borough Council and Newcastle City Council around smart energy planning. After an initial scoping exercise, ESC provided dedicated resources to work with each area to develop Smart Energy Plans for each area. These plans were envisaged to set out the themes that each area wished to focus on in developing future innovation projects/demonstrators linked to the ideas of SSH.

As part of the grant for SSH2, ESC committed to deliver both an interim and final evaluation of the programme.

Methodology

Evaluation formed part of the contracted deliverables of the programme. This included: the development of a detailed Evaluation Approach (in the form of an Evaluation Plan); a supporting Data Collection Plan; an Interim Evaluation Report; and a Final Evaluation Report.

To meet this requirement, when the SSH2 programme was originally funded, ESC planned to run a large and comprehensive evaluation alongside the programme, comprising in-house and contracted resource. However, delays and recruitment difficulties (ESC were unable to identify and recruit a suitably skilled embedded evaluation manager for the programme) meant that a more modest approach needed to be undertaken. As a result, Winning Moves (an external consultancy) were commissioned to support ESC's delivery of the evaluation.

As introduced above, the SSH2 programme was funded between Summer 2017 and March 2019. Winning Moves were appointed to support the evaluation in Winter 2017 (at which point ESC had drafted the Evaluation Plan and Data Collection Plan internally). Winning Moves supported the finalisation of the Evaluation Plan and, on behalf of ESC, led the interim evaluation (with a report in June 2018) and the final evaluation (with a report in March 2019).

The aims of the evaluation and key evaluation questions were agreed in Winter 2017, with a final evaluation approach agreed by Spring 2018.

The approach to the evaluation was theory based. The programme demonstrated a number of attributes which can be set-out in a theory of change. These included multiple and diverse 'interventions' and projects; the likely impacts being long term and the intended impacts being difficult to measure (possibly intangible) (Stern, 2015).

Although there are obstacles in implementing theory based evaluation (Weiss, 1997a and 1997b) it was concluded that, in the context of SSH2, a theory based design could help to deal with this diversity and the long term nature of the potential impacts. It was envisaged that a theory of change approach would help to get to grips with this complexity, by specifying and testing the contexts, drivers and barriers (referred to as assumptions) that bring about outcomes. In addition, SSH2 is trialing new activities, for which the understanding of the causal mechanisms was not yet known or established. Again, it was felt that a theory of change approach could help to articulate and test this; opening up the 'black box' to answer not simple the question of what works, but also why and how it worked (Carter, 2012, Chen, 1990, Stame, 2004). Lastly, a theory of change approach would examine the role SSH2 played in bringing about the observed results, thereby creating learning.

The following sources were used to provide evidence to assess the evaluation questions and to test and refine the theory of change:

Qualitative interviews with internal stakeholders. Interviews were conducted with a purposive selection of ESC staff involved in delivering SSH2, and employees of BEIS, SSH2 project sponsors. The interviews were conducted face-to-face where feasible (by telephone when not) using a qualitative topic guide to instigate an open discussion between participant and researcher. Broadly, the topics covered in the evaluation interviews with internal stakeholders covered SSH2 delivery, engagement with stakeholders and learning from the programme. Interviews were recorded, transcribed and key points recorded in a spreadsheet (alongside supporting quotes).

Qualitative interviews with external stakeholders. Interviews were conducted with a selection of external stakeholders. The sample population was compiled and provided by ESC and comprised any organisations that ESC had attempted to engage with as part of the programme. A sampling approach was proposed on the basis of organisation type (manufacturers / hardware developers, local authorities, energy service providers, policy / regulators and registered social landlords) and the extent of ESC interaction (high or low). Sub-contractors on the programme were also included in the sample. Interviews were conducted by telephone using qualitative questions to instigate an open discussion between participant and researcher. Eight of the external stakeholder interviews were conducted with partners involved in delivering the programme (including those on the steering group for the work). The topics covered in the evaluation interviews with external stakeholders covered awareness and understanding of SSH2 and interest and engagement with SSH2 outcomes. The questions varied by organisation type in line with the potential outcomes that could be achieved by each group. Key points from each interview were recorded in a spreadsheet (alongside supporting quotes). Interviews were recorded to aid recall of key points.

Analysis and review of programme deliverables and programme delivery documents. Through the course of delivering SSH2, ESC produced a variety of documents including information documents (mainly for an internal audience) and programme deliverables. ESC also collated information provided by third parties as part of delivering the programme, including information about stakeholders and input from third parties. As part of governance structures, a number of project management documents were also produced and updated. Corporate communications were also produced and circulated. A number of these programme deliverables and data items were reviewed. The key points from each report were recorded in a spreadsheet, alongside researcher assessment of the implications for the evaluation.

The analysis of all the evidence was organised by the theory of change and the evaluation questions using a structured analytical framework[1]. The approach (Framework analysis) sought to find evidence to exemplify the theory of change in line with the approach to analysis of qualitative data set out by Gibbs (Gibbs, 2007). An analytical framework was used to structure the data which was summarised and reduced from the different evidence sources in order to support answering the evaluation questions and testing and informing the theory of change. Charmaz suggests some basic questions are used in Framework analysis that move the analysis beyond the descriptive to analytical. The final question was particularly pertinent for the analysis in this setting 'How do structure and context serve to support, maintain, impede or change these actions and statements?' (Charmaz, 2003) as this asks the evaluator to think about the role of the 'system' in bringing about the actions and changes suggested by the evidence.

The framework was organised in a spreadsheet format with each row representing an evidence requirement or question (e.g. there was a row for each output identified in the theory of change) and each column a sub-set of the data (e.g. the views of staff at BEIS formed a column in the spreadsheet). Evidence was summarised in the most appropriate cell (along with the source of that evidence). Evidence synthesis was done two ways, in sequence, after a workshop to ensure shared understanding the evidence requirements and further familiarisation time. Method 1 involved systematic review of the individual spreadsheets for each evaluation element to identify relevant evidence. Method 2 involved reflecting on the key information required and seeking evidence for this based on familiarity with fieldwork undertaken. Different researchers were responsible for review of different sub-sets of the data and their synthesis into the analytical framework. For verification purposes, all content was reviewed and challenged by a second researcher.

Direct quotes were added to the spreadsheets, taken from transcriptions of recorded interviews or directly from reports and documents where they exemplified a particular piece of evidence. Some of these have been used in the report representing the most exemplary quote for the point that is being made and a spread of respondents (i.e. to avoid relying too much on a few participants).

[1] This approach was adapted from the Framework Method most commonly used for the thematic analysis of semi-structured interview transcripts

When the analytical framework was complete, and based on all the sources of evidence, a judgement was made about the strength of the evidence supporting different elements of the theory of change and a code was introduced for this – weak, medium and strong. The judgement was made based on both the breadth of the evidence (e.g. was it arising from one or multiple sources) and the depth of the evidence (e.g. the level of detail and examples). This reflects the joint judgement of the evaluation team.

Results

Although SSH2 was funded as, and referenced as, a programme, the evaluation approach needed to go beyond the features of a programme evaluation. A programme evaluation typically looks at a single intervention on an organisation, cluster of organisations, or sector. The evaluation of SSH2 was required to explore a portfolio of actions; there were individual interventions within three broad categories of activity. The evaluation was also seeking to understand the potential impact of this portfolio of activity a system; the energy system. Evaluation does not get easier as we move from the project and programme levels towards considering sub-systems and systems (Arnold, 2004). Reflecting on how systems innovation is defined by the community (systemsinnovation.io) the challenges in evaluating the system presents themselves in different ways:

- A complex system is typically a large-scale system composed of many interdependent parts that are relatively autonomous. Therefore, the evaluator needs to identify all the parts of the system and identify how their 'movement' can be understood.
- Exploring the innovation of this complex system, is about understanding whether the parts are organised, connecting or communicating in a new and valuable way. The evaluator needs to understand how this will manifest itself, what would be tangible evidence of the potential for this.

Tactics for dealing with some of the difficulties in evaluating systems include investing time to scope the evaluation (what can reasonably be evidenced) and making explicit use of theory as a 'benchmark' against which to judge whether innovation can or is emerging (Arnold, 2004). A further tactic is the use of methods to reach judgements at aggregated levels i.e. whether the sum of the parts can or will be different, not just the parts. All were employed on the evaluation of SSH2.

Through delivering the evaluation, evaluation stakeholders built understanding of what has and will work with regards to systems innovation and the methods and approaches that do, don't and might be used to address this. This paper will first set out our learning about what did work about how to evaluate systems innovation, before describing issues that remained unresolved through the evaluation.

What did we learn about how to evaluate systems innovation?

The aims and purpose of the evaluation need to be clearly scoped and framed. When developing an evaluation plan, having a clear idea about the aims of the evaluation and the questions that need to be addressed by an evaluation, and for whom, is crucial to inform the design of the evaluation and the expertise required (Mayne and Stern, 2013).

This is additionally important within the context of a programme designed to innovate a system, which is about developing learning in its own right. Therefore, it needs to be considered how the learning developed through the evaluation is distinct from the learning developed through the programme itself (so the evaluation is doing more than simply describing the programme learning) (moving the analysis from descriptive to analytical, Gibb, 2007). Some of the activities within the SSH2 programme included evaluative elements in-of themselves. In particular, the field trial which was looking to develop insight into consumer preferences for Heat as a Service, preferences for low carbon alternatives and understanding of how to deliver Heat as a Service (i.e. to determine what works with regards to delivering Heat as a Service). There was a risk that the evaluation could overlap with

some of these activities, and it was important for the evaluation work to build on these, as opposed to repeating programme learning.

The conversations and considerations when scoping the aims and questions for evaluation of SSH2 included the anticipated audience for the evaluation, their requirements and how they would use the outputs. It took approximately two months for these conversations to conclude. There were challenges in balancing a desire to 'sweat' the evaluation for all that could be learnt against the pragmatic evaluator perspective to provide a realistic scope to what could be evidenced. There was also a challenge in developing a set of questions that were relevant for the life-time of the programme. They needed to be at a sufficiently aggregated level to accommodate the various ways in which the programme might adapt over the course of delivery, without risking evidence that was over-generalised or missed the detail. For example, making sure we captured what the programme had delivered, whilst acknowledging that the exact detail of this might change over the course of delivering the programme.

The primary users of the findings of the evaluation were the Energy Systems Catapult. This included senior management within the organisation, as well as those responsible for delivering this phase and future phases of the SSH programme. It was identified that they would use the evaluation within the timescales of the evaluation, to inform decision making aimed at programme improvement; to contribute to the broader evidence base for business case development and to support future funding applications.

BEIS were also a key evaluation audience. As funders of the SSH2 programme, it was identified that they would look to the evaluation to demonstrate accountability for the programme (in line with programme deliverables) as well as to contribute towards the broader evidence base for policy development. It was also acknowledged that BEIS have extensive evaluation expertise and would provide methodological input to the evaluation, to align the work with other BEIS evaluation work on innovation programmes.

Finally, the Energy Systems Catapult has a number of wider programme stakeholders and organisations that they were seeking to influence through the programme. Whilst they were not an intended audience for the evaluation, it was felt that it should be borne in mind that the Energy Systems Catapult may seek to use evidence from the evaluation in their efforts to engage this audience.

With this in mind, the overarching aim of the evaluation was to provide understanding and evidence that identifies, supports and improves pathways to decarbonise domestic heat in cost effective ways in order to meet agreed carbon targets by 2050.

A secondary aim was to demonstrate accountability by providing understanding and evidence of the role that the SSH2 programme has played in delivering this.

On this basis, in Winter 2017, the core aims of the evaluation were agreed as follows:

1. 'Programme learning' aim: To provide evidence of what has been delivered alongside understanding that supports and improves SSH2 delivery (evidence that can inform steering and implementation of the programme).
2. 'System learning' aim: To provide understanding and evidence of the extent to which SSH2 is developing an effective pathway to decarbonisation.
3. 'Strategic learning' aim: Reflecting beyond SSH2, to provide broader understanding and evidence about the models, capabilities, and contextual factors required to enable market innovations to be delivered at sufficient scale to meet 2050 targets.

Do not underestimate the challenges in drafting a theory of change. As part of this evaluation, a theory of change was developed at the outset of the evaluation in consultation with SSH2 programme staff.

The process of developing the theory of change shared a number of known features of this approach. It involved ESC and the evaluators working collectively and iteratively, starting with the aims of the programme, and seeking to identify a plausible and testable set of actions that could result in those aims being met (Connell Kubisch, Schorr and Weiss, 1995; Fulbright, Kubisch and Anderson, 1998).

The theory of change set out the expected pathways towards the overall outcomes and goal of SSH2 through describing the activities, outputs and expected outcomes of the programme. It was a useful way of understanding, making explicit, and defining the assumed connections (i.e. the theory) between the activities undertaken as part of the programme and their anticipated impacts.

There were a number of benefits of the process in which the theory of change was developed, including the instigation of exploratory conversations with programme staff (Allen and Bicket, 2018). This was useful for the evaluators to understand some of the emergent thinking behind the programme and the way in which delivery was evolving. The process also ensured that programme managers were involved in key decisions about which, of the many possibly elements of the theory, became the focus for the actual evaluation activity (Blamey and Mackenzie, 2007) e.g. the final report includes a table reporting against each of the outputs as agreed in the theory of change.

The theory of change was reviewed at the conclusion of the interim evaluation and a number of revisions proposed. These informed fieldwork activities for the final evaluation.

A more in-depth review of the theory of change was conducted during the final evaluation, involving further consultation with SSH2 programme staff. The final theory of change encapsulates shared understanding of how SSH2 activities worked as understood at the conclusion of the programme and the key assumptions behind them. Due to the timescales in which these revisions were made, the final theory of change did not inform fieldwork activities for the final evaluation. However, it was used to structure the analysis.

The final theory of change was significantly different to the one developed at the outset of the evaluation both in terms of content and representation; the theory of change developed at the outset of the evaluation was largely in narrative form, whereas the final theory of change was visualised. The significance of the amendments reflected increased understanding of what the programme could do (within the context of the wider decarbonisation agenda) as well as an evolving understanding of theory of change as a tool to structure evaluation learnings.

Although there are a wide variety of ways of diagrammatically representing a theory of change (Funnel and Rogers, 2011), including to evaluate a system (Rogers, 2017) the theory of change for SSH2 was depicted as a results chain or pipeline logic model. This approach could be argued to be somewhat reductionist in its understanding and presentation of how the activities within SSH2 were seeking to bring about systems innovation. Other evaluators have depicted systems in a less linear and more dynamic way (Howe, 2015; Barbrook-Johnson, Penn, Kaxira and Ahmed, 2019) However, the purpose of the theory of change in the evaluation of SSH2 was to try and bring some structure to what was being delivered, in order to test this more systematically.

The effort expended to agree a final theory of change and the fact that the theory of change evolved substantially in content and presentation over the course of the evaluation was largely down to two key challenges to engaging with theory of change as a tool.

Within the early stages of programme delivery, Energy Systems Catapult developed a Benefits Management Framework. This was a theoretical framework of expected benefits of SSH as a whole setting out how the programme will ultimately transform the market that enables the required wide decarbonisation of domestic heat. The Benefits Management Framework was not used to identify, define, track, realise and optimize benefits in the traditional sense of a Benefits Management approach (Jenner, 2012) but was used as a communications tool. It was used both internally and externally to rationalise and explain the programme and its intended benefits. The Benefits Management Framework used some of the language of theory of change (specifically 'outputs' and 'outcomes') and was therefore seen as offering some of the functionality of the theory of change. This apparent 'overlap', plus the extent to which the Benefits Management Framework was embedded into early delivery of the programme meant it was challenging to find a space and purpose for a theory of change alongside this.

Secondly, the whole suite of work of the SSH2 programme was building understanding and evidence of how to decarbonise heat and exploring to what extent this can be achieved by changing the market to one which delivers heat as a service. This was a long term objective, which was not (and was not intended to be) achieved

within the timescales of the SSH2 programme. In consultation with SSH2 programme staff, there were challenges in committing to the outcomes of the programme (sufficient to be able to succinctly document them within a theory of change) when these were not anticipated to be delivered within the timescales of the programme and were widely acknowledged to be influenced by factors outside of the programme. Within the context of a wide set of influencing factors (such as market, regulatory and policy drivers) there were challenges in agreeing what outcomes might be and the relative role of ESC in contributing to these outcomes.

In drawing conclusions from the evaluation, it was subsequently recognised that theory of change could have been a useful tool in programme delivery. If the theory of change had evolved alongside the Benefits Management Framework they might have been seen as complementary tools. Whereas the sequencing of their development and the fact that external stakeholders built familiarity with the Benefits Management Framework, meant that the theory of change was effectively in its shadow. One of the learnings of the evaluation with regards to the management of SSH2 was that there needed to be a focus on decarbonisation – and how this is being delivered – running through all project activity**.** A key objective of SSH2 was to reveal and test an approach to decarbonisation of UK housing stock. However, it was not clear to all stakeholders that there was a strong link between the activities carried out within the programme and the decarbonisation objective. This link was also missing from the original programme narrative. In this regard, tools, such as theory of change, could have been used at an earlier stage in the programme to articulate the link to low carbon heating.

The value of building in informal feedback loops. There were two points at which evaluation evidence was synthesised and reported. Early findings from the evaluation were used to provide interim findings to BEIS in June 2018 with a final evaluation report providing overall insight and learning required to feed into end of project reporting in March 2019. The short time frame between these two reports supported 'faster' learning from the programme. The interim report set up expectations of what would be explored in the final evaluation.

In addition to the more 'formal' feedback structures, interim findings were shared informally with ESC in a discursive feedback session, prior to formal report drafting. This informal feedback provided an opportunity to embed learning within the programme team approximately four weeks earlier than waiting for the formal report (useful in the context of a relatively short programme). Activities to inform interim findings focussed on clarifying the programme theory, and answering the first evaluation question for the study through a review of programme documentation, interviews with the programme delivery team, external delivery partners and external policy stakeholders. Interim evidence did not however feed into the benefits management strategy (as suggested by UK Government Evaluation guidance as set out in the Magenta Book), which would have been an additional way to embed rapid feedback into the programme, allowing corrective action to be taken (Infrastructure and Projects Authority, 2017; Jenner, 2014) and more closely align the theory of change with the Benefits Management Framework.

The interim report confirmed that the programme was broadly on track to deliver. It made early recommendations about the way dissemination could be better tailored to key stakeholder groups and communicating how the Heat as a Service (Haas) concept is linked to decarbonisation. Additional learnings highlighted through the discursive feedback session, highlighted the following, which all elicited a response from ESC:

1. The evidence surfaced the opportunities to improve dissemination activities with external policy stakeholders. It was acknowledged by both the ESC programme team and staff at BEIS directly involved in SSH2 that dissemination activity with policy stakeholders had not been optimal prior to Spring 2018, and that this was an area that needed to be addressed. Having acknowledged the issue, both programme leads and BEIS were talking about how it could best be addressed. Evaluation findings contributed to efforts to change the frequency and nature of engagements with external policy stakeholders
2. The evidence highlighted the need to bring the programme team together and ensure a shared vision of what the programme should be aiming to achieve within the remainder of the funded period. Lots of programme activity prior to Spring 2018 had necessarily been focused on successful delivery of the

winter trial and maximising the data and insight this generated. Consultations that informed the interim evaluation identified that it was a good time to bring the programme team together and ensure a shared vision of what the programme should be aiming to achieve within the remainder of the funded period. Wider programme aims were broadly understood, it was the specific aims that need to be (re)articulated and disseminated. For example, what were the outcomes that the programme is seeking from dissemination activity with each stakeholder group? Evaluating findings contributed to the decision to schedule an organisation-wide session to discuss this issue.

3. There was evidence that throughout the course of SSH2 delivery, partners had high expectations of the outcomes of SSH2. This was, in the main, a legacy of their involvement in SSH1. Based on conversations in SSH1, delivery partners hoped for a larger scale of trial than was delivered in SSH2 plus they expected to have responsibility for part of the delivery budget in SSH2 and to have some delegated delivery responsibility. However, the reality of SSH2, was that delivery partners performed more of a consultative role, being consulted on decision making and kept informed of the progress of the programme and outcomes. Discussions at the interim evaluation stage reminded the team about the ongoing benefits of active management of external delivery partners and affirmed that this was an area where continual investment was required.

What issues and challenges about how to evaluate systems innovation remained unresolved through the evaluation?

Timing of the final evaluation. The evaluation was conducted concurrent to delivery of SSH2 as the deadline for the final evaluation was the same as the date for programme conclusion (as a result of the Final Evaluation report being a deliverable for the SSH2 programme).

In addition to this being a contracted requirement of the programme, there were some operational advantages of this (compared to situations when evaluation activity continues post-programme conclusion). Programme stakeholders, and in particular programme staff, were able to devote time to evaluation activities as they were still being employed to deliver programme activities; this benefitted the scheduling of interviews as well as the provision of data and programme documentation.

However, the deadline for the final evaluation report meant that fieldwork activity to inform the evaluation was conducted in January and February 2019. This impacted on the data available to inform the evaluation. The timing limited the extent to which the achievement of intended outcomes (especially downstream policy changes and market effects that may not yet be apparent) have been captured by the evaluation. For example, at the time of fieldwork it was too early to see whether the Smart Energy Plans developed as part of the local engagement work were delivering on their potential for local area heat decarbonisation and / or inspire activity at local level in other areas.

In particular, the timing limited the ability of the evaluation to capture market reaction to the learning from the programme e.g. how Energy Service Providers, potential / actual hardware providers (including hub providers) responded to the lessons and learnings from the programme. The interviews with external stakeholders to inform the evaluation were undertaken concurrent with final dissemination activities. In particular there were two well-publicised public outputs from the work that sought to share insight and learning with Energy Service Providers (ESP)and wider industry – a final report from the work and a final showcase event. It was envisaged that ESPs and wider industry would use the data, insight and learning arising from the trial to help them design interoperable heat products, services and policies / regulations, to support and deliver decarbonisation of heat through Heat as a Service. In order to encourage use of the data, insight and learning arising from the trial, ESC actively disseminated this. This was on the basis of a Dissemination Plan detailing how they were planning to communicate findings from the programme including through events and conferences, publishing articles, and information shared through the ESC website and social media channels. There were also plans to make the data available.

How to engage with programme stakeholders in an appropriate and timely way. As set out above, a key part of the evaluation methodology involved engaging with what were termed 'external' programme stakeholders i.e. industry and local government organisations and individuals who the programme were seeking to influence through the learning. The evaluation sought to explore the following for key stakeholders:

- Awareness of the SSH2 programme vision and goals
- Understanding of the SSH2 programme goals, including their [the stakeholder's] role in delivering them, and what would need to change to achieve them. This included seeking feedback on SSH2 activities, whether they are delivering what is needed, and if so whether these are delivering effectively
- Desire for action to change and evidence of real action, working towards the goals.

The challenge for the evaluation was that stakeholder groups adapted through the programme and the messaging of the programme evolved during delivery. To some extent this was a function of the evolving nature of the programme as an innovation programme. It might be expected that the messaging of the programme evolves through testing and refinement with stakeholders. However, this required an evaluation approach that was capable of adaptation and a dynamic and responsive approach to interviewing. Stakeholders had different levels of engagement with the programme and for several of them, a range of issues were relevant in interviews.

For example, the original evaluation plan included an intention to interview hardware providers. The Benefit Management Framework for the SSH2 programme identified hardware providers as a target audience they wanted to influence through the SSH2 programme. The Benefits Management Framework set out an expectation that, following the outcomes of the SSH2 programme, it was anticipated that there would be evidence of low carbon product vendors developing products to work within an open architecture for retailers to integrate (e.g. heat pump vendors, battery vendors etc.). However, through the course of programme delivery, in particular the learning through the business modelling work, this expectation evolved. It was envisaged that a future gateway (HESG) might be a cloud-based platform which would interact primarily with proprietary in-home hub devices, which would in turn interact with the hardware, rather than an ESC HESG in the home directly interacting with hardware such as heat pumps or valves. As a result, stakeholders adapted to include potential hub providers in addition to hardware providers.

With regards to messaging, there was a lack of a strong low carbon narrative to support communicating with stakeholders; the value of the work in revealing and substantiating paths to low carbon was therefore diminished in the view of those stakeholders. The link to decarbonisation was not clearly evident in the way that the programme was explained by the programme team to those outside of the programme. It was not well understood by external stakeholders how decoupling heating outcomes (warmth, comfort) from energy purchased (kWh) provides a pathway to domestic decarbonisation or how offering domestic consumers 'Heat as a Service' overcomes barriers to moving to low carbon energy sources at both pace and scale. Approaching this concept in interviews with external stakeholders required a careful and responsive approach to reflect the different levels of understanding of, and perspectives on, this link.

Value of conducting the evaluation externally vs internally. As set out above, when the SSH2 programme was originally funded, ESC planned to run a large and comprehensive evaluation alongside the programme, comprising in-house and contracted resource.

In Winter 2017, ESC ran a recruitment campaign for an Evaluation Lead. The vision was that in the first 12 – 15 months of the role, they would be expected to deliver an interim and final evaluation of the second phase of the SSH programme. Beyond this, it was anticipated that the Evaluation Lead would be involved in building in evaluation to any future phases of the SSH programme and other Catapult activity. It was also anticipated that the role would help ESC to embed an evaluation-led culture in their project designs, helping them to build their internal skills and capabilities through knowledge-transfer whilst being the 'go to' person for evaluation questions.

The role of Winning Moves was to deliver the Evaluation Lead role for the SSH2 programme, until an internal appointment was made. Thus, the intention was not to develop a dependency on an external consultancy to deliver the work, but for an external consultancy to:

- Assist the ESC in development of evaluation capacity and capability
- Provide external capacity and capability (as the ESC's internal provision develops)
- Offer evaluation expertise through maintaining a critical friend role.

However, delays and recruitment difficulties (ESC were unable to identify and recruit a sufficiently skilled evaluation manager for the programme) meant that a more modest approach needed to be undertaken. As a result, Winning Moves were commissioned to support ESC's delivery of the evaluation.

There were a number of advantages to involving an external organisation in the delivery of the programme evaluation:

- Winning Moves were able to bring objectivity to the exercise. In particular, internal stakeholders took value from the interviews conducted to inform both the internal and final evaluation. The objective nature of the questions / discussion in the interviews sparked thoughts for respondents.
- Use of an external organisation, which necessitated the agreement of a budget, resources and timetable, brought a discipline to the evaluation exercise. This was quite useful in the context of the evaluation being delivered concurrent to programme delivery, where there might have been rationale to delay decisions or outputs (to capture further delivery). Decisions were required, for example about who should be interviewed (in the context of an agreed number of interviews).

However, there were some also some challenges in involving an external organisation that may have been less pronounced with a dependency on internal resource only:

- Remaining informed about current project delivery. The SSH2 programme was a reasonably large programme with multiple delivery activities within three workstreams. As an innovation programme, programme delivery was constantly being adapted in response to learning. It was difficult for an external organisation, who was not immersed within the delivery environment, to keep up to date with programme delivery. This may have limited the extent to which the evaluators were able to uncover the programme theory and the reality of programme delivery (Blamey and Mackenzie, 2007).
- Further informal feedback and making connections. Related to the above, there were undoubtedly opportunities to embed evaluation learning within aspects of programme activity that were missed due to the team delivering the evaluation not being co-located with hose delivering the programme and surrounded by day to day project delivery.

Conclusions

To achieve the target of net zero greenhouse gas emissions from across the UK economy by 2050, the UK must accelerate the rate and scale at which the domestic heating market is decarbonised. Given the rate and scale of change required we need to learn fast about what can work (and won't work) to understand how to achieve this transition in the energy market. Evaluators can play a key role in supporting innovation programmes to deliver this learning, but this does not come without its challenges; and to a certain extent, compromises.

The experience of evaluating the SSH2 programme suggested that the evaluation of programmes designed to inform or achieve systems innovation should observe all of the well-understood features of a well-prepared, planned evaluation. The key features of the evaluation process include defining the audience for the evaluation, identifying the evaluation objectives and research questions, selecting the evaluation approach,

identifying the data requirements, identifying the necessary resources and governance arrangements, conducting the evaluation and using and disseminating the evaluation findings (HM Treasury, Magenta Book).

The application of a theory-based approach worked well for this evaluation. In the context of a programme designed to inform how to change a system, the theory of change itself was instrumental in scoping what could be plausibly measured by the evaluation, particularly in terms of outputs and outcomes. The process of developing the theory of change through workshops, conversation and iteration was also valuable in terms of developing the evaluators understanding of the programme, the system and the key actors within the system. The process of developing the theory of change also facilitated buy-in to what the final evaluation report would report on.

Arguably, the theory of change would have been more useful if developed earlier in the programme and used as a programme management tool. However, it would not have had the same value for the evaluation if this had not been developed with the evaluators. The process of programme managers and evaluators working on the theory of change together ultimately meant that the evaluators were 'closer' to the programme and more able to feed in valuable insights on an informal basis (allowing programme managers to learn 'faster'). These rapid insights would have added more value if they had been part of a benefits management approach; in which case the insights could have been targeted at immediate benefits realisation. This suggests evaluations of other initiatives seeking to understand how to innovate a system might benefit from a holistic approach to monitoring, benefits management and evaluation.

Reflecting more widely, the feature of the evaluation that delivered particular value in the context of SSH2 were taking the time at the start to identify the evaluation objectives and research questions; to get organisational buy-in to these and to position the evaluation as sitting alongside the programme; observing and contributing, but not duplicating effort. Establishing informal mechanisms for disseminating evaluation learnings also delivered value, enabling a more rapid response to learnings from the evaluation than waiting for, and digesting, a more formal report.

Although the experience did not conclude whether or not it is more advantageous for an evaluator of an innovation programme to sit within or outside the policy or programme delivery organisation, it did identify the way in which the evaluator needs to operate to deliver the evaluation effectively. The evaluator needs to be highly responsive to the dynamic delivery environment; willing to make changes to timescales and approach as well as on the fly adaptations in gathering evidence, particularly in primary research interviews. The evaluator needs to be comfortable with uncertainty and potentially incomplete information; evaluating concurrent to delivery there is rarely an ideal time for evaluation activity and, where evaluation activity is delivered it is likely that that the evaluator is not in full possession of all the background or context. Lastly, the evaluator needs to work to keep implementers engaged and bought into the evaluation; continually demonstrating its value and minimising the what may be perceived as the disruptive consequences of evaluation activity.

References

Allen, Richard and Bicket, Martha 2018. Complexity, policy and evaluation in practice: Improving theories of change to address complexity. Policy Evaluation for a Complex World.

Arnold, Erik 2004. Evaluating research and innovation policy: a systems world needs systems evaluations. *Research Evaluation*, Volume 13, Issue 1, April 2004, 3–17.

Blamey, A and Mackenzie, M 2007. Theories of change and realistic evaluation: Peas in a pod or apples and oranges. Evaluation, 13(4), 439–455.

Barbrook-Johnson Peter, Penn Alex, Kaxira Anna, Ahmed Tajbee, 2019. Negotiating complexity in evaluation planning: A participatory systems map of the energy trilemma. CECAN Evaluation and Policy Practice Note (EPPN)

No. 12 for policy analysts and evaluators., In: CECAN Evaluation and Policy Practice Note (EPPN) for policy analysts and evaluators Centre for the Evaluation of Complexity Across the Nexus (CECAN)

Infrastructure and Projects Authority 2017, *Guide for Effective Benefits Management in Major Projects.*

Carter, R 2012 Governance and Social Development Resource Centre, University of Birmingham.

Charmaz, K. 2003. 'Grounded Theory', in J.A. Smith (ed.), *Qualitative Psychology: A Practical Guide to Research Methods*. London: Sage.

Chen, H.T. (1990) *Theory-Driven Evaluations*, Thousand Oaks, CA: Sage

Connell, J. and A. Kubisch (1998) 'Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems', in K. Fulbright-Anderson, A. Kubisch and J. Connell (eds) *New Approaches to Evaluating Community Initiatives*, vol. 2, *Theory, Measurement, and Analysis.* Washington, DC: Aspen Institute.

Connell, J. P., A. C. Kubisch, L. B. Schorr and C. H. Weiss (1995) *New Approaches to Evaluating Community Initiatives*, vol. 1, *Concepts, Methods and Contexts*. Washington, DC: Aspen Institute.

Fulbright-Anderson, K., A. Kubisch and J. Connell, eds (1998) *New Approaches to Evaluating Community Initiatives*, vol. 2, *Theory, Measurement, and Analysis.* Washington, DC: Aspen Institute.

Funnell, Sue C and Rogers, Patricia J 2011. Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.

Gibbs, G. R., 2007. Thematic coding and categorizing. *Analyzing Qualitative Data*. London: SAGE Publications, Ltd.

Howe, Catherine 2015. Systems Thinking: Theory of Change in Curious Catherine Blog.

HM Treasury, 2020. *Magenta Book*.

Jenner, S 2012. Managing Benefits. APMG International.

Jenner, S. 2014. *Managing Benefits.* 2nd Edition. The Stationery Office.

Mayne J. and Stern E. 2013. Impact evaluation of natural resource management research programs: a broader view. ACIAR Impact Assessment Series Report No. 84. Australian Centre for International

Rogers, Patricia 2017. Better Evaluation FAQ: How do you use program theory for evaluating systems? Blog for Better Evaluation.

Stame, Nicola 2004. Theory-Based Evaluation and Types of Complexity. Evaluation. January: 58 – 76.

Stern, E, 2015. Impact Evaluation: A Guide for Commissioners and Evaluators. Prepared for the Big Lottery Fund, Bond, Comic Relief and the Department for International Development.

http://systemsinnovation.io/ [Accessed 2nd June 2020]

Weiss, Carole H 1997. How can theory based evaluation make better headway?. Evaluation Review, August: 501-524.

Weiss, Carole H 1997. Theory-based evaluation: Past, present, and future. New directions for evaluation.

Winning Moves. *Smart Systems and Heat Phase 2 D32: Evaluation Report*. 2020