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

# Can energy projects be over evaluated?

*Sam Hampton[1] & Tina Fawcett[1]*
*[1]Environmental Change Institute, University of Oxford, United Kingdom*
*samuel.hampton@ouce.ox.ac.uk*

## ABSTRACT (250 WORDS)

The UK government is investing £102.5m into a programme called 'Prospering from the Energy Revolution' (PFER). PFER includes four energy 'demonstrator' projects, and 10 'design' projects. A consultancy will evaluate the entire PFER programme; a consortium of researchers spanning 22 universities (EnergyRev) and the Energy Revolution Integration Service will evaluate aspects of the projects; which will also be closely monitored by the funder (Innovate UK). This paper focuses on one of the demonstrator projects, 'Energy Superhub Oxford' (ESO), on which a *fifth* team of evaluators is conducting further evaluation work. Identifying a gap in the energy evaluation literature, this paper introduces the idea that projects can be *over-evaluated.* It seeks to identify what this might look like, its implications, and how it can be managed. Findings from the first year of PFER show that there is significant overlap in the aims and scope of each evaluator, with potential for duplication, wasted resources and consultation fatigue. ESO partners feel overwhelmed by requests for data and information; are struggling to prioritise; and are confused about the role of each evaluator. We define three characteristics of over-evaluation: *too many cooks, too much consultation,* and *too much data,* and discuss their potential positive implications as well as their more obvious downsides. We argue that energy transition trends may make over-evaluation more common, and avoiding its pitfalls may not be straightforward. As the first paper of its kind, we aim to stimulate debate amongst the energy evaluation community about the prevalence and dangers of over-evaluation.

### Introduction

For decades, the energy evaluation community has been campaigning for the greater use of evaluation for energy projects and programmes. The main challenges faced by this community are extensively discussed in the literature. Many energy related projects are subject to no formal evaluations at all (Flett et al., 2018; NAO, 2013). Where they are incorporated into project budgets, they may be subject to various flaws including constrained resources, inadequate theories of change, self-reporting bias, lack of independence, or a focus on impact evaluation to the detriment of process or market evaluation (Vine et al., 2012a; Johansson et al., 2017).

In all the discussion of challenges, pitfalls, theoretical models and best practice approaches, there has been no documented discussion in the energy evaluation literature of the dangers of *over-evaluating* policy-based initiatives. This can be explained in large part because it is more common to have insufficient resources than too many, but it is possible that when over-evaluation does occur, we do not fully understand how to recognise it, or how to deal with it. It is also possible that certain trends in the energy system make it more likely for over-evaluation to occur in future. Demand flexibility, smart-technologies, the closer integration of electricity, heating and cooling and transport and the shift towards decentralised generation each mean that energy systems are involving an increasingly wide set of actors, from users to technology providers, to public funding bodies operating at various levels of government. Demonstration projects are becoming increasingly complex, involving more stakeholders and audiences, and requiring multi-dimensional evaluations.

Set against this background, this paper argues that it is possible to over-evaluate energy projects. Justified by the dangers of wasted budget and resources, and overburdening project stakeholders, participants and beneficiaries, it calls for greater attention to be paid to over-evaluation, and provides a first attempt at critically analysing its occurrence in publicly funded energy projects. The paper is guided by two main research questions:

(1) What are the characteristics of over-evaluation, and how can it be identified?

(2) What are the positive and negative implications of over-evaluation, and how can these be managed?

Our evidence is drawn from the UK, where there has been a steady stream of funding for 'energy demonstration projects' in recent years, as different options are pursued for ensuring the energy transition is environmentally sustainable, affordable and rapid. We use as a case study the £102.5m (€123m) 'Prospering From the Energy Revolution' (PFER) programme which includes four energy 'demonstration' projects, and 11 'design' projects. The programme as a whole includes funding for four separate bodies with an explicit evaluation remit, in addition to those involved in evaluation work at the project level. This paper focuses one of the demonstrator projects, called 'Energy Superhub Oxford' (ESO), which is deploying grid-scale battery storage, coupled with initiatives to electrify domestic heating and accelerate the uptake of electric vehicles in Oxford. Starting in early 2019 and running to 2022, our evidence is derived from interviews, documentary analysis and direct participation in the first year of the project. We intend to follow up with a reflective paper for EEE 2022.

Following a brief review of literature, we introduce the case study and introduce the five evaluators working on the Energy Superhub Project, highlighting their different (and overlapping) roles and evaluation approaches. The following section presents findings from interviews with project stakeholders and documentary evidence. The discussion outlines the three main features of over-evaluation, and discusses the challenges of avoiding their pitfalls while maximising their positive potential. We conclude by calling for greater attention from the energy evaluation community to over-evaluation, which may become more prevalent as the energy transition accelerates.

## Literature Review

As an area which spans research and practice, the energy evaluation literature is made up of a combination of academic peer-reviewed publications and extensive grey literature, including guides and briefings produced or commissioned by governmental bodies. The majority of this literature falls into three categories.

Firstly, there are a vast number of papers which present results and learnings from specific projects and programmes, many of which are comprehensively archived at energy-evaluation.org. Besides reporting on impacts, outcomes, implementation processes and providing policy recommendations, this body of literature often draws on individual case studies to contribute wider insights for the evaluation community. These observations and recommendations span a range of themes, including how to increase skills and capacities amongst the evaluation community (Vine et al., 2012b; Vine, 2019); how multiple benefits can be estimated (Killip et al., 2019; Ryan and Campbell, 2014); and how evaluation of energy efficiency projects might be expanded to include wellbeing metrics in their scope (Campbell, 2019).

A second category of energy evaluation literature are those analyses which review multiple project evaluations, with an aim of making generalisations for audiences of project commissioners and practitioners (Broc et al., 2018; Dougherty and Van de Grift, 2016; Harmelink et al., 2008; NAO, 2013; Sandin et al., 2017; Sheate et al., 2016; Vine et al., 2012a). These include audits of swathes of government-funded programmes, which often report on a lack of consistency in approach and objectives, and a worrying absence of evaluation in many cases. An example from the UK is the National Audit Office's cross-cutting review of publicly funded projects, in which it found significant variation in the budgets, approaches and impacts of evaluations when commissioned by different government departments. Not only did it find that a minority of evaluations provided sufficient evidence of policy impact, it also stated that 'government fails to use effectively the learning from these evaluations to improve impact and cost-effectiveness' (NAO, 2013). In 2018, the UK Energy Research Centre conducted a review of evidence from 119 energy system demonstration projects funded since 2008 (Flett et al., 2018). It identified a dearth of reporting from these projects, with most outcomes self-reported by project consortia, and 'only a limited number subject to independent evaluation'. These meta-reviews of evaluation provide valuable insights for funders and practitioners, identifying common mistakes and barriers to be overcome (Broc et al., 2018; Sandin et al., 2017; Vine et al., 2012a), and scaling up results beyond single projects (Bukarica and Tomšić, 2017; Flett et al., 2018; Kushler et al., 2012). However, no examples could be found which report the existence of over-evaluation.

A third category of literature sets out to provide guidance and advice for evaluation best practice, and is largely made up of reports written or commissioned by government departments and agencies (HM Treasury, 2011; Kushler et al., 2012; Li et al., 2018; State and Local Energy Efficiency Action Network, 2012). Such is the scale of this literature, that the US based Consortium for Energy Efficiency even published a 'guide to the guides' (2008). Li et al's (2018) report on the US based Uniform Methods Project, for example, runs to over 1000 pages, made up of detailed evaluation guidance covering energy applications from variable frequency drives to whole-building retrofit; while the State and Local Energy Efficiency Action Network published guidance for evaluating energy efficiency programmes at the state and utility scale (2012). Despite a wealth of advice for impact, process and market evaluations provided by this extensive literature, no discussion could be found of the dangers and pitfalls of over-evaluating energy projects.

These three categories of energy evaluation literature share two characteristics which are relevant to this paper. Firstly, there is no discussion of, or concern for, the existence of over-evaluation in publicly funded energy projects, nor any guidance for avoiding its pitfalls. There are two possible explanations for this. On the one hand, it is possible that over-evaluation is an extremely rare occurrence, and as such has not been reported on in the literature. On the other hand, it is possible that when it does occur, policy professionals, academics or practitioners have failed to identify it as such, or have reported on its characteristics in other ways. Survey fatigue, for example may be explained by other factors; or perhaps the over-collection of data has not reported on to avoid embarrassment.

The second commonality across energy evaluation literature is an assumption – often implicit - that evaluation work is conducted by a single party. While there is ample discussion of the status, role and responsibility of evaluators, whether they are internal, arms-length or fully independent of the funding and delivery bodies, there has been no consideration of cases where multiple organisations share responsibility for evaluating energy projects. As the next section demonstrates, this assumption is no longer sufficient, and there are trends associated with the energy transition which indicate that multi-stakeholder evaluation may become more common in future.

## Prospering From the Energy Revolution

Incorporating £102.5m (€123m) of government investment and leveraging more than three times this from private sector funding, the PFER programme consists of three parts. There are four 'demonstrator' projects, 11 energy 'design' projects, and a range of 'research and integration services' involving academic researchers and policy analysts. The demonstrator projects make up the bulk of the expenditure, implementing innovative energy technologies across heat, transport and electricity vectors, and embedded in local areas. Two projects are based in Oxfordshire, one in the Orkney Islands off the north coast of Scotland, and another in West Sussex.

The focus of this paper is Energy Superhub Oxford (ESO) which is being implemented by a consortium consisting of four private sector technology developers, Oxford City Council and Oxford University. The central aim of the project is to mitigate the need for costly upgrades to the constrained electricity distribution network in the city, by opening up a new 60MW connection to the UK transmission network. A 50MW/55MWh hybrid battery, using lithium-ion and vanadium flow technologies in parallel, will be installed at Oxford's main transmission sub-station, providing grid balancing services and minimising the need to curtail variable sources of generation. The remaining 10MW capacity will be used to provide electric vehicle charging services to businesses and consumers in the city, using a dedicated cable which will run alongside Oxford's ring road to rapid-charging hubs. Finally, the project will also see the installation of over 300 ground-source heat pumps in social housing, coupled with smart controls which optimise heating schedules based on variable prices of electricity.

The PFER programme as a whole includes funding for four separate bodies with an explicit evaluation remit. Two independent consultancies, Ipsos Mori and Technopolis will work together to evaluate the entire PFER programme. A consortium of 60 energy researchers spanning 22 universities, called EnergyRev, will evaluate aspects of the demonstrators and design projects. Finally, the Energy Revolution Integration Service (ERIS), which is part of a government-funded agency called the Energy Systems Catapult, will provide additional insights using primary and secondary data and modelling techniques. In addition to these three organisations, which have remits spanning the entire PFER programme, each of the demonstrator projects are closely monitored by the funding body, Innovate UK. Furthermore, the projects have capacity within their consortia for further analysis of impacts and procedures.

Our small research team of two social scientists and two engineers fall into this *fifth* category of evaluators. When putting together the proposal for funding ESO, it was not clear what capacity there would be for evaluation across the PFER programme. The private sector company leading the bid approached our team, and we designed a package of work to evaluate the environmental, social and economic impacts of the project, as well as conducting in-depth modelling and analysis of the technical performance of the hybrid battery. The project was awarded funding by the science and innovation body Innovate UK and began in April 2019. Table 1 describes the evaluators' roles and approaches taken as they relate to the ESO project.

This paper is informed by 10 months' direct involvement in Energy Superhub Oxford, as a consortium member. Over the early stages of the project, we have conducted interviews with four senior individuals from partner organisations, as well as one senior civil servant from Innovate UK. Publications and briefing documents are used to map out the intersecting scopes of the various evaluators and their different approaches.

| **Organisation** | **Innovate UK** | **Ipsos Mori and Technopolis** | **EnergyREV** | **Energy Revolution Integration Service (ERIS),** | **Oxford Univ. Environmental Change Institute and Dept of Engineering** |
|---|---|---|---|---|---|
| **Organisation type** | Science and innovation funding body | Private sector consultancies | Academic consortium, including 60 researchers across 22 universities | Publicly funded think tank, part of Energy Systems Catapult (ESC) | Academic |
| **Function and remit** | Funder and monitoring function | Overall programme evaluation | Research spanning systems modelling to policy, business and user focus | Helping to overcome barriers and create insights | Social and environmental impacts (ECI); battery performance (DoE) |
| **Specialism and approach** | Performance monitoring and project governance | Impact, economic and process evaluation | Specialist skills including modelling, policy and governance, user engagement, business models and systems thinking | Markets and consumers | Local knowledge and expertise across energy and transport; Battery modelling |

| | | | | | |
|---|---|---|---|---|---|
| **Funding arrangements** | Direct grants from central government | Awarded the PFER Programme evaluation on a competitive basis, reporting to Innovate and BEIS. | Created as part of the PFER programme, reporting to Innovate UK. | ESC funded publicly and through offering private services. Report to Innovate UK, and BEIS | Funded through ESO project, reporting to partners and Innovate UK. |

## Early findings from ESO and the PFER programme

This section presents findings from direct experience of working on ESO and the PFER programme, as well as interviews with stakeholders from funding bodies and partners. Given the small number of contributors, direct quotations are avoided, and instead their views are summarised.

One respondent from Innovate UK explained the rationale behind investing so many resources into monitoring, evaluation and research. When joining 5 years ago, they had inherited a number of smaller scale technology projects which had been funded without plans for independent evaluation, and they indicated that while projects could claim to have been successful according to a variety of metrics, their overall contribution to the energy transition was unclear. Recently, they had been working hard to build in more evaluation into their funded projects, and that on the PFER programme an opportunity arose to work more closely with UK Research & Innovation, the body responsible for the majority of public research investment in the UK. This included the opportunity to produce quantitative insights for the energy system alongside the conventional focus on business growth and technical innovation.

Unlike other domains of innovation, energy projects face a unique challenge due to their complexity. Whereas when they funded innovation on other topics, such as to boost the production of off-site, modular housing in the UK, they were able to easily communicate the aims and strategic approaches adopted to audiences from policy makers to business and the taxpayer. The PFER programme on the other hand, is investigating the viability and scalability of 'smart, local energy systems', incorporating innovative technologies and business models across electricity, heating and transport. The initial proposal for funding had been rejected by the government department which funds Innovate UK, requiring 'hundreds' of redrafts in order to convince Ministers of its value. Projects which can succinctly explain what benefits they will achieve, and how, tend to get more money and more resources. Once the PFER programme had been funded, communicating its objectives and its impacts became a priority for Innovate UK, and an additional rationale for appointing several evaluators who could help to articulate these.

Despite 2 years of planning, the launch of the PFER programme was somewhat hasty, as senior Ministers urged its roll-out. This meant starting multiple, individually complex initiatives in parallel, stretching resources and exposing insufficient processes. The consequence was that little consideration was given to how projects and objectives may overlap and could complement one another. As a result, the precise scope and objectives of each of the evaluators has had to be worked out during the first stages of the programme and new information is still being generated and shared at the time of writing, 10 months after the projects began. Unsurprisingly, as the evaluators are sharing their plans with one another, significant areas of overlap have emerged.

For example, the question of whether the demonstrator projects can be *scaled up and reproduced* in other parts of the UK and internationally is common to the evaluation frameworks published by Ipsos Mori, EnergyRev, ERIS and Oxford University. Ipsos Mori identify a key question in their extensive, 116-paged, evaluation framework: 'to what extent did the Challenge lead to investability, scalability and replicability of new models created?', while ERIS have identified scalability and replicability as two of their seven key evaluation criteria. EnergyRev include scaling up as one of their six 'key thematic areas', in which they intend to both analyse issues such as drivers and barriers, unanticipated and negative effects of replication, as well as providing support to projects through workshops and guidance to overcome these challenges. Finally, our team of scientists at Oxford University developed eight objectives, which included 'developing insights into how the demonstration project may be successfully reproduced in other cities across the UK'. Beneath these overarching objectives, each evaluator offers some detail about how the methods of data collection they intend to use to address the question of replicability. Inevitably, these overlap, as each intend to interview project partners, survey users and supply-chain stakeholders, and conduct desk-based analysis of the social, economic and political characteristics of the geographically bounded areas in which the demonstrators are based.

Another example is the focus on analysing *policy implications.* Another of EnergyRev's six themes is devoted to 'policy, regulation, markets, governance', and the academic consortium not only plans to undertake a 'systematic review to compile an evidence map of the current policy and regulatory landscape', but also to convene a 'Policy Contact Group', which will provide a bridge between academics and businesses working on the PFER programme and a range of policy decision makers. Similarly, Ipsos Mori aims to 'evaluate the extent to which learnings from [PFER] are utilised and inform investments, policies and regulations over time', while Oxford University includes an aim to develop 'insights for policy, governance and regulation for the transition to smart, local energy systems' as part of its eight objectives.

Despite these major overlaps, there are several elements of the evaluators' work which appear to complement one another. ERIS, for example, have chosen to 'set' their evaluation in 2032, adopting an imagined retrospective analysis of the demonstrators' performance 10 years after they conclude. They intend to conduct detailed calculations and modelling of the effects and potential effects of the projects on the unit price of energy, as well as the 'network costs', both of which are anticipated to fall as a result of the introduction of flexible demand, machine-learning trading algorithms, and increased storage capacity. They expect to go further than any other evaluator in conducting a detailed calculation of the greenhouse gas emissions reductions which result from the demonstrator projects.

Similarly, there are aspects of EnergyRev's work which will not be covered by other evaluators. This includes their 'cyber-physical systems' theme, which will go beyond the traditional remit of an evaluator by developing 'optimal distributed control and operation algorithms', which will be trained on project data to inform the operation and governance of smart-grids in the future. Oxford University's Engineering department also has ambitions which reach beyond the scope of conventional evaluation. They will create a 'digital twin' of the grid-scale hybrid battery, in order to simulate the degradation of lithium-ion under a variety of conditions. Their model will produce insights about the value of combining lithium-ion and vanadium flow technologies as a means of extending battery-life, and their unique research expertise provides a contribution to the PFER programme that could be offered by no other evaluator.

As explained above, the different evaluators' plans for primary data collection are still being finalised, and the nature of overlaps are being discussed as a matter of urgent concern. Nonetheless, several requests for data, information and other input have already been made to ESO consortium partners by each of the five evaluators. This includes a request from Innovate UK to partners to estimate a wide range of expected impacts resulting from the project, from 2018 to 2027. The request included more than 120 individual impact measures, ranging from 'renewable energy generated (in MWh)', to 'average energy price reduction per unit (p/kWh)'. As a result, some of the partners already described the effects of *consultation fatigue.* One partner explained how they were required to report numbers in different ways to a variety of external organisations. Having been cooperative at first, they were now taking a more relaxed attitude to requests for data.

Two consortium partners expressed frustration with the fact that the evaluators had taken more than six months to clearly articulate their approach and what they could offer the projects. One explained that as well as overlapping interests, evaluators were working at different speeds; some still consulting on methodologies, and others already requesting data. Another was frustrated by the fact that evaluators had not begun consulting at an earlier stage, and that some of the evaluators' approaches remained vague. Feeling overwhelmed and somewhat confused by the various documents being produced and requests coming from the evaluators, one project partner struggled to see the value in the numerous data requests. However, when explaining to evaluators that they were unable to respond quickly or fully to the requests, had received sympathetic responses.

Having developed a relationship with the Oxford University researchers through regular project meetings, one partner that they preferred to communicate with this in-project team, and would refer other evaluators to them in future, rather than respond directly.

Another partner expressed concerns that as evaluators were beginning to 'come alive', this inevitably meant there would be more 'calls on our time'. This partner explained that they were 'worried' and 'nervous' about the demands being placed on both the project managers and the consortium partners, and hoped that having the Oxford University team within the consortium would mean that some of the requests could be funnelled through this internal evaluation team, to avoid taking up others' time.

In summary, findings from the first year of the PFER programme indicates that there has been a degree of duplication of effort and arguably, over-allocation of evaluation resources. For project consortium partners, these have resulted in some confusion and consultation fatigue, and there remains a need for the five evaluators to coordinate their work with more care as the programme continues. In light of these findings, the next section outlines the characteristics of over-evaluation, and describes some positive consequences besides the more obvious negative implications.

## The characteristics of over-evaluation: downsides and upsides

Two features of over-evaluation emerged from analysis of early findings from the PFER programme, and a third is an anticipated challenge as the programme goes on. The first is *too many cooks*, the second is *too much consultation,* and the third is *too much data* (Figure 1).

*Too many cooks* is shorthand for the English adage 'too many cooks spoil the broth', which describes the effects of having too many people participate in a task, resulting in suboptimal outcomes. This phenomenon is distinct from the supposed problem of over-allocating resources to a single evaluator. While there may be evidence of this being a challenge for some energy projects, the administration of a large budget by a single evaluator is likely to be easier to manage than the division of resources between multiple organisations with different priorities, methodological approaches and audiences.

The problem of *too many cooks* may lead to a range of negative outcomes. Without clear and deliberative planning prior to, or during project initiation, there is a danger of different evaluators designing their work with overlapping scopes. This may result in each of them targeting the same groups of respondents for surveys and interviews, with the risk of duplication of effort, wasted resources, and confusion on behalf of respondents. If finding themselves in this situation, as the PFER evaluators do, 10 months into the programme, there is a need for regular communication and consultation in order to minimise these negative outcomes. However, changing scope and objectives midway through a project can be problematic, particularly where these are written into contractual agreements and linked to project key performance indicators (KPIs).

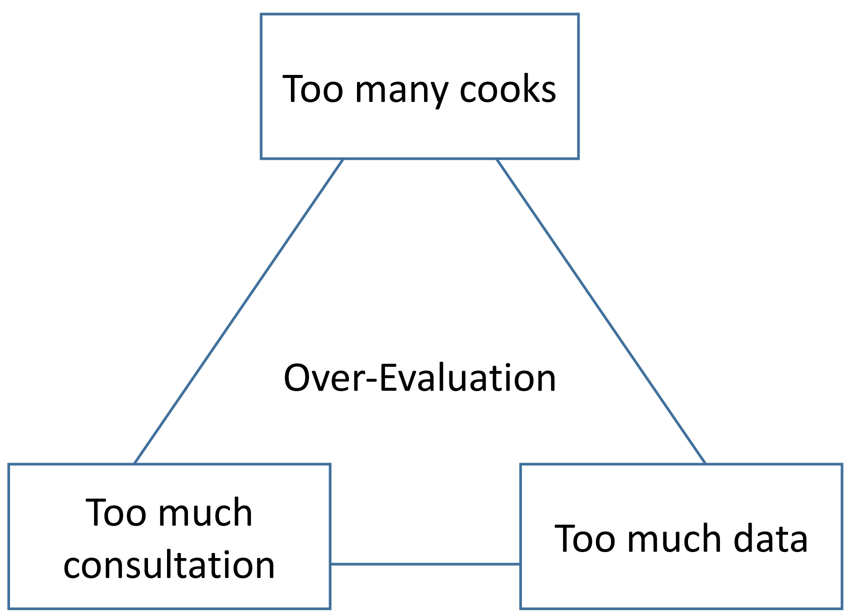

*Figure 1: Three features of over-evaluation*

Despite these significant dangers, there are potential upsides to the *too many cooks* scenario. For instance, multiple evaluators can offer diverse perspectives on energy innovation projects. Where these, like ESO, span multiple energy vectors and involve stakeholders from government to large and small businesses, transport users and social housing tenants, there is logic behind applying the diverse expertise of consultancies, universities and think-tanks to evaluation. Whereas in many documented cases of energy evaluations, evaluators tend to focus on particular impacts (Sandin et al., 2017), or might employ a limited range of methods such as economic cost-benefit analysis (Harmelink et al., 2008), having multiple evaluators may help to ensure that some of the more frequently neglected aspects of evaluation, such as user experience or policy analysis, are given adequate attention. On the PFER programme, the funders' hope was that by having diverse and multiple perspectives would mean that the whole is more than the sum of its parts. Despite the evidence of overlapping scope described in the previous section, findings also highlighted complementary approaches adopted by the different evaluators.

As well as adopting diverse perspectives, multiple evaluators can also appeal to a variety of audiences, maximising the reach of project learnings. On PFER, Ipsos Mori will be adopting an overarching view of the programme, reporting to central government funders on questions such as cost effectiveness. They are adopting a global outlook, as they seek evidence of the UK's leadership in developing smart, local and integrated energy systems. As academics, EnergyRev and the Oxford University team will aim to public outputs in peer-reviewed journals, for an international audience of energy scholars. The contributor from Innovate UK recognised the value of academic outputs for their greater reach and longevity. They indicated that while White Papers could be more focused and timely, they tended to generate little international interest, and would be soon forgotten as the policy cycle moved on.

A third upside to *too many cooks* which is emerging on the PFER programme is *agility* on behalf of the evaluators. On major 'living laboratory' projects such as ESO, changes in scope, timeframes and unanticipated obstacles are an inevitability (Castán Broto and Bulkeley, 2013). ESO has experienced each of these, discovering, for example, the existence of protected woodland at the battery site, and complications associated with its underground cable crossing municipality boundaries. A good evaluation must be able to adapt its focus to meet the changing scope of the project itself, whilst also documenting and reflecting on the meandering path of implementation. On PFER, without clearly delineated responsibilities at the outset, the various evaluators have demonstrated flexibility with one another, and are adapting the changing circumstances of projects themselves.

Finally, the high degree of communication and collaboration that having *too many cooks* requires may also lead to higher overall standards. Sharing their methodologies and objectives with one another at draft stages, the evaluators on the PFER programme have been able to shape and respond to each other's plans. Like academic peer-review, these processes are not always easy for the parties involved, but ultimately lead to improved outcomes.

The second problem associated with over-evaluation is *too much consultation.* The main negative effect of this is 'consultation fatigue' (Murray et al., 2009; Hess et al., 2012), which can take multiple forms with various implications. When overwhelmed by requests to complete surveys, for example, response rates can fall (Sinickas, 2007), or requests for interviews are rejected by project stakeholders such as business leaders, policy makers and funders on the basis that they have already given evidence elsewhere. The contributor from Innovate UK suggested this was a common problem in the energy sector, where consultations were issued frequently. Besides low response rates, they indicated that this leads to bias, whereby large incumbent energy companies with an interest in lobbying are able to devote greater resources to consultations. Smaller companies, with fewer resources, struggle to get their voices heard.

There are few upsides to the problem of *too much consultation*. However, one example, if respondents are willing to complete more than one set of interviews or surveys, is that results may be compared for verification of results. Ample evaluation resource may allow for longitudinal data collection, or can be channelled into boosting response rates. In any case, the preferred approach involves combining efforts to maximise coverage while minimising the number of requests made to respondents.

The third characteristic of over-evaluation is associated with generating *too much data.* Although the PFER programme and its demonstrator projects are in their early stages and have so far produced little quantitative data, there are indications that this may become a concern at a later stage. For example, responding to a request from the Innovate UK monitoring team for projects to identify KPIs, the ESO consortium produced a total of 179 indicators. These include simple metrics such as 'number of charges per day' by electric vehicles, to those which will involve *millions* of data-points such as humidity readings in over 300 households with heat-pumps, and minute-by-minute electrolyte temperatures for each module of the vanadium flow battery.

The danger of having *too much data* is that it becomes unwieldy to manage and analyse. On ESO, the wide variety of data produced during the project will require careful and meticulous record-keeping by project managers, who will need to ensure that data is collected, archived at regular intervals, is represented in consistent format, and is shared with relevant stakeholders while ensuring compliance with data protection regulations. Extracting value from vast quantities of data also demands significant capacity and expertise. Experience from previous demonstration projects, as well as the nature of academic publishing indicates that researchers are likely to analyse *some* data in great depth, for specific outputs. The Oxford University team, for example, anticipate that their work will utilise only a small proportion of the total data produced in the ESO project.

It may be argued that you cannot have *too much data*: that there is no harm in retaining information whose value may be not be immediately apparent, but may be extracted at a later date. Besides the resources required for data management, this may be true. However, experience of analysing research data suggests that interpretation of evidence is often most pertinent and most nuanced when the researcher participates in data collection and holds a stake in the project. Qualitative methods literature, for example, has highlighted the value in researchers transcribing their own interviews (Davidson, 2009; Oliver et al., 2005). When processing and analysing data that one has had a hand in collecting, its *resonance* is retained, and the process guides the insights and narratives that are generated. Quantitative data may not hold the same significance for the researcher, but its meaning and significance is likely to be enhanced when combined with qualitative methods and analysed by individuals close to the project. The nature of smart, local energy projects, incorporating a wide variety of technologies stakeholders, demand both mixed-methods approaches, and synthesising skills on behalf of evaluators.

## Dealing with over-evaluation

One way to avoid the pitfalls of over-evaluation on energy innovation projects is to appoint a single evaluator with a broad remit and extensive skills and capacity. This is the incumbent model for energy projects, as our review of three categories of evaluation literature testifies. However, as the energy transition becomes more urgent and ambitious, cross-cutting programmes such as PFER are likely to become more commonplace, testing the abilities of even the largest consultancies and think-tanks. This approach would also exclude the specialist expertise that exists in smaller evaluation consultancies. While requiring coordination and oversight, a multi-agency approach has many advantages, and is already widely used in evaluating international responses to humanitarian emergencies (Brusset et al., 2010). Trends outlined in this paper indicate that the single evaluator model may become a rarer occurrence in energy system innovation too.

Another recommendation is that when multiple evaluators are involved, clear remits and responsibilities should be set out at the start to avoid overlap and confusion. However, this is not always possible in practice due to the nature of innovation funding. The launch of PFER was accelerated at the request of government ministers, and projects proposals were submitted without knowledge of how the programme as a whole would be evaluated. The ESO proposal writers included Oxford University researchers as internal evaluators, only to find later that they could draw expertise from the EnergyRev consortium or ERIS. As discussed in the previous section, this confusion has led to some positive outcomes, as the different evaluators have been reviewing one another's frameworks, adapting scope to reflect expertise and capacity, and exercising flexibility which will benefit the programme as a whole.

Once projects involving multiple evaluators are underway, effective working requires clear and regular communications; an open approach to data sharing; and a spirit of cooperation and transparency. These can be difficult to achieve however. Evaluators will be guided by different objectives, audiences, and the geographical scales on which they focus. When spanning the public and private sector such as on ESO, evaluators are likely to operate using different linguistic norms and theoretical approaches. Consultancies for example, may have objectives and deliverables written into contracts, restricting flexibility, and may be reluctant to share proprietary information such as methodologies or existing datasets. Academic researchers' motives can sometimes be opaque to funding bodies. For instance, while they are likely to be driven to produce peer-reviewed publications for an international audience, disciplinary traditions often dictate they use idiosyncratic language and conventions, rendering outputs accessible only to a narrow group of specialists. Sharing data also has its challenges. Data protection regulations and ethical guidelines for research data may prevent free exchange, while the process of data analysis benefits from an intimacy with the data itself, which can be lost when those analysing have had no hand in methodology design, or fail to appreciate its context.

Aiming to avoid the situation of having *too much data,* project partners would ideally develop a selective, goal-oriented data collection strategy, including a detailed specification of scheduling, storage, and sharing. This process may be most effectively conducted towards the end of the project-initiation phase, as partners and evaluators each begin to focus on delivering their work packages, and detail on the sources, types and quantities of data become increasingly known. Following this model, the ESO consortium agreed their KPIs 6 months into the project. However, having identified 179 different fields, it may be argued that they have not adopted a *selective,* nor *goal-oriented* approach to data collection. Perhaps driven by ambition; naivety; responding to pressure from various evaluators; and without knowledge of the precise future uses of data, project managers and partners will now need to dedicate significant time and resource to maintaining good records. It remains to be seen whether such ambitious plans for data collection will be produce beneficial or problematic outcomes.

This section has provided a set of recommendations for avoiding the pitfalls of over-evaluation. Some of these, such as those involving planning, open communications and consensus agreements, may be considered rather obvious. Implementing these may less than straightforward, however, and we have highlighted the conditions under which the features of over-evaluation may occur and be difficult to overcome.

## Conclusions

This paper has introduced the idea of over-evaluation into the energy research literature for the first time. Drawing on evidence from UK's Prospering From the Energy Revolution programme, we have outlined three main features of over-evaluation (1) *too many cooks,* (2) *too much consultation, and* (3) *too much data.* Focusing on the first year of the PFER programme and Energy Superhub Oxford, which includes *five* separate evaluators, we have found evidence of both negative and positive consequences of having *too many cooks, too much consultation,* and have identified the risk of producing *too much data* as the project goes on. Negative consequences include duplication of effort, wasted resources, and confusion and consultation fatigue on behalf of respondents. Vast quantities of data require extensive management resource, and can be wasted without sufficient analysis. However, we also identified several potential features which may offset these downsides. These include incorporating diverse perspectives and expertise; appealing to a wider range of audiences; and encouraging agility and high standards from evaluators. Well-maintained data archives may be mined at a later date for purposes as yet unknown. This outline of the nature of over-evaluation is intended to stimulate debate amongst the energy evaluation community, and has potential use as a framework for identifying over-evaluation in projects within and beyond energy.

Historically, it has been far more common that energy projects face the problem of being *under*-evaluated than over-evaluated. Nonetheless, we consider it unlikely that the evidence presented in this paper is unprecedented. We argue that over-evaluation warrants attention from researchers, practitioners and funders. It is more likely that the occurrence of over-evaluation has not been documented, rather than being non-existent. It may be argued that it is not be in the interests of evaluators to report on the phenomenon, and given that a large proportion of this body of work is produced by evaluators themselves, this may explain the gap in the literature. Another explanation is simply that what we have defined as over-evaluation, others may define as the effects of poor project management and lack of oversight from funding bodies. Our discussion of how to avoid the pitfalls of over-evaluation suggests that this explanation is insufficient however, as there are a variety of structural forces which make dealing with the challenge of over-evaluation more than a question of resource management.

There are reasons to believe that the circumstances discussed in this paper may become more common in future, as the energy transition becomes more urgent and more ambitious. The UK's PFER programme has funded four major demonstration projects with a combined value of over £150m (€177m), and involves a number of public institutions from central government to research bodies and innovation agencies with diverse interests and expectations. The demonstration projects span multiple energy-vectors and involve stakeholders ranging from local government to private business, electricity regulators, householders and transport users. Evaluators must have diverse expertise and extensive capacity to engage with all aspects of these cross-cutting projects, as well as the ability to disseminate results and learnings to audiences spanning local residents to international investors. As programmes such as PFER continue to be funded in response to the climate crisis and the need for rapid energy system transformation, the limits of the conventional single-evaluator model will be exposed. In future, the features of over-evaluation outlined in this paper may become more prevalent, demanding further research and debate.

This is an exploratory paper, based on evidence from the early-stages of the PFER programme. It has raised a deliberatively provocative question and seeks to generate debate. We intend to follow up with a further paper at the Energy Evaluation Europe conference in 2022, when we will have had the chance to review the arguments put forward here in light of complete evidence from the PFER programme, which ends in March 2022. In the meantime, the questions raised would benefit from comment and evidence from the energy evaluation community.

## References


Broc, J.-S., Thenius, G., di Santo, D., Scholmann, B., ver der Meulen, J., Maric, L., 2018. What can we learn from sharing experience about evaluation practices? Presented at the International Energy Program Evaluation Conference (IEPEC), Vienna, Austria.

Brusset, E., Cosgrave, J., MacDonald, W., 2010. Real-time evaluation in humanitarian emergencies. New Directions for Evaluation 2010, 9–20. https://doi.org/10.1002/ev.326

Bukarica, V., Tomšić, Ž., 2017. Energy efficiency policy evaluation by moving from techno-economic towards whole society perspective on energy efficiency market. Renewable and Sustainable Energy Reviews 70, 968–975. https://doi.org/10.1016/j.rser.2016.12.002

Campbell, N., 2019. Energy efficiency and wellbeing benefits – a New Zealand approach. Presented at the Energy Evaluation Asia Pacific Conference, Bangkok, Thailand.

Castán Broto, V., Bulkeley, H., 2013. Maintaining Climate Change Experiments: Urban Political Ecology and the Everyday Reconfiguration of Urban Infrastructure. International Journal of Urban and Regional Research 37, 1934–1948. https://doi.org/10.1111/1468-2427.12050

Consortium for Energy Efficiency, 2008. Energy Efficiency Program Evaluation: A Guide to the Guides.

Davidson, C., 2009. Transcription: Imperatives for Qualitative Research. International Journal of Qualitative Methods 8, 35–52. https://doi.org/10.1177/160940690900800206

Dougherty, A., Van de Grift, S.C., 2016. Behavioral energy feedback program evaluations: a survey of current knowledge and a call to action. Energy Efficiency 9, 899–909. https://doi.org/10.1007/s12053-016-9445-8

Flett, G., Kelly, N., McGhee, R., 2018. Review of UK Energy System Demonstrators. UK Energy Research Centre.

Harmelink, M., Nilsson, L., Harmsen, R., 2008. Theory-based policy evaluation of 20 energy efficiency instruments. Energy Efficiency 1, 131–148. https://doi.org/10.1007/s12053-008-9007-9

Hess, S., Hensher, D.A., Daly, A., 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: Policy and Practice 46, 626–644. https://doi.org/10.1016/j.tra.2011.11.008

HM Treasury, 2011. The Magenta Book: Guidance for evaluation.

Johansson, I., Stenqvist, C., Thollander, P., 2017. Energy Efficiency Networks for SMEs - Program Theory and Ongoing Evaluation. Presented at the 2017 IEPPEC Asia-Pacific Conference on Evaluation of Energy Efficiency Programs and Policies, 1-7 November, Bangkok, Thailand.

Killip, G., Fawcett, T., Cooremans, C., Crijns-Graus, W., Krishnan, S., Voswinkel, F., 2019. Multiple benefits of energy efficiency at the firm level: a literature review, in: ECEEE Summer Study Proceedings. Presented at the ECEEE, Belambra, Presquile de Giens, France.

Kushler, M., Nowak, S., Witte, P., 2012. A national survey of state policies and practices for the evaluation of ratepayer-funded energy efficiency programs (No. U122). ACEEE.

Li, M., Haeri, H., Reynolds, A., 2018. The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures (No. NREL/SR-7A40-70472). National Renewable Energy Laboratory, Golden, CO.

Murray, M., Fagan, G.H., McCusker, P., 2009. Measuring horizontal governance: a review of public consultation by the Northern Ireland government between 2000 and 2004 [WWW Document]. https://doi.org/info:doi/10.1332/030557309X462556

NAO, 2013. Evaluation in Government. National Audit Office.

Oliver, D.G., Serovich, J.M., Mason, T.L., 2005. Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research. Social Forces 84, 1273–1289. https://doi.org/10.1353/sof.2006.0023

Ryan, L., Campbell, N., 2014. Estimating the Macroeconomic Impacts of Energy Efficiency Policies and Measures. Presented at the International Energy Program Evaluation Conference (IEPEC), Berlin, Germany.

Sandin, S., Neij, L., Mickwitz, P., 2017. Review of evaluations of policy instruments for energy efficiency in buildings in Sweden, in: ECEEE Summer Study Proceedings. Presented at the ECEEE, France.

Sheate, W.R., Twigger-Ross, C., Papadopoulou, L., Sadauskis, R., White, O., Orr, P., Phillips, P., Eales, R., 2016. Learning lessons for evaluating complexity at the nexus: a meta-evaluation of CEP projects (Final Report to CECAN).

Sinickas, A., 2007. Finding a Cure for Survey Fatigue. Strategic Communication Management; Chicago 11, 11.

State and Local Energy Efficiency Action Network, 2012. Energy Efficiency Program Impact Evaluation Guide, Evaluation, Measurement, and Verification Working Group. Prepared by Steven R. Schiller, Schiller Consulting, Inc.

Vine, E., 2019. Creating an Evaluation Community in Asia: What's Happened in the Last Two Years?, in: Energy Evaluation. Presented at the Energy Evaluation Asia Pacific Conference, Bangkok, Thailand.

Vine, E., Hall, N., Keating, K.M., Kushler, M., Prahl, R., 2012a. Emerging issues in the evaluation of energy-efficiency programs: the US experience. Energy Efficiency 5, 5–17. https://doi.org/10.1007/s12053-010-9101-7

Vine, E., Saxonis, W., Peters, J., Tannenbaum, B., Wirtshafter, B., 2012b. Training the Next Generation of Energy Efficiency Evaluators: A Report from the Field. Presented at the International Energy Program Evaluation Conference (IEPEC), Rome, Italy.