An Experimental Research Design for Evaluating Energy Feedback

BEHAVE 2016
4th European Conference on Behaviour and Energy Efficiency
Coimbra, 8-9 September 2016

AN EXPERIMENTAL RESEARCH DESIGN FOR EVALUATING ENERGY FEEDBACK

Martin Pullinger[1]*, Nigel Goddard[2] and Janette Webb[3]

1: School of Informatics
University of Edinburgh
10 Crichton Street, Edinburgh, UK, EH8 9AB
e-mail: martin.pullinger@ed.ac.uk, web: http://edin.ac/1T6WheV

2: Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh
10 Crichton Street, Edinburgh, UK, EH8 9AB
e-mail: nigel.goddard@ed.ac.uk, web: http://homepages.inf.ed.ac.uk/ngoddard

3: School of Social and Political Science
University of Edinburgh
Chisholm House, High School Yards, Edinburgh, UK, EH1 1LZ
e-mail: jan.webb@ed.ac.uk, web: http://www.sps.ed.ac.uk/staff/sociology/webb_janette

Keywords: Experimental research design, Mixed methods research, Behaviour change interventions, Energy feedback

Abstract

The role of digital energy feedback technologies in encouraging and supporting behaviour change is the subject of substantial current research effort. Effective evaluation of such technologies is an essential but challenging process for generating new knowledge in the field about what effects interventions have, and why. This paper describes the research design for the IDEAL project, a UK EPSRC-funded project developing and testing an innovative system to provide domestic energy feedback, tailored to the energy practices of each household, via a tablet computer. The paper highlights our experiences with the complexities of designing experimental interventions tested in real-world settings (i.e. actual households) over extended periods of time. We focus on those aspects of design which offer potential insights for similar projects, current and future. Our central conclusion is that for this kind of study, it is necessary to investigate the co-production of technological components and social processes in addition to measuring specific changes in domestic energy use.

1. INTRODUCTION

This paper focuses on the principles and practicalities of undertaking a multi-stage experimental evaluation of a home energy monitoring and feedback system. It describes our experiences of this for the IDEAL project, a major UK EPSRC-funded project running from 2013-2017, which is developing and evaluating an innovative system to monitor domestic energy-using practices, and provide energy feedback, tailored to the energy practices of each household, via a tablet computer. The project's main hypothesis is that 'a personalised behavioural feedback loop is more effective in inducing demand-reducing behaviour change than [a] consumption feedback loop' [1]; that is, feedback that focuses on and is tailored to the energy using *practices* of each household will be more effective than conventional home energy feedback that focuses purely on patterns of energy *use* and associated financial and carbon costs. We comprise a multi-disciplinary team of social scientists, wireless sensor network and machine learning researchers, database and software systems engineers, and HCI researchers from the University of Edinburgh. The project aims to make contributions across these fields to the understanding of energy demand, practices and feedback design, to inform future policy and system design.

The purpose of this paper is to draw out learnings from the process of developing and implementing IDEAL's research design that can be of use to inform research methods in other experimental studies of energy demand management systems. Despite substantial research in this field, relatively few studies have focused on experimental trials of such systems (although see, for example, [2],[3]), and there is little published on applying experimental methods 'in the wild' for such research, that is, in real home settings. Section 2 briefly describes the main considerations in developing the project's research design, many of which will be similar to those facing other large scale evaluation projects: the research aims, principles of research design, existing literature in the relevant disciplines, and the practical considerations involved in real-world research. Section 3 describes the resultant research design that we are implementing, from study design to data collection and analysis. Section 4 discusses potential issues and risks when undertaking such complex evaluation studies, and section 5 concludes by summarising the key considerations for sound research design, and its role in ensuring that the evidence base is as robust as possible.

2. KEY CONSIDERATIONS OF A MULTI-STAGE EXPERIMENTAL RESEARCH DESIGN

2.1. Defining and achieving the project aims

Clear aims and objectives are important to guide all aspects of work in a project, including how the research design should be structured. IDEAL, reflecting its multi-disciplinary team, has various interlinked research aims covering different areas of the project:

1. Sensor network: The initial stage was to develop and test a low power wireless sensor network capable of gathering robust data on a home's gas and electricity use, and room temperature, light and humidity levels, at a resolution of seconds, and transmitting the

data to our secure server.

2. Energy disaggregation methods: A machine learning (ML) module aims to develop and test the methods required to draw on these data streams, along with data on the home's appliances and occupants, to infer the performance of energy using practices in the home, e.g. frequency, duration and timing of showering or heating patterns, and to link these to their impacts in terms of energy use.
3. Energy feedback design: This module aims to develop and evaluate an enhanced energy feedback user interface that provides personalised feedback on a household's energy using practices and their impacts in terms of energy use, conditions in the home, etc. Feedback aims to draw on the sensor data and ML inferences to tailor the feedback delivered to each home in a rule-based manner (e.g. by delivering advice for practice changes that is most relevant to a particular household).
4. Understanding energy using practices, their impacts, and their 'drivers': This module aims to test different social science disciplinary perspectives on understanding factors shaping energy use and practices, using the mixed methods data collected in the study; and to explore implications for the wider smart cities, sustainability and energy demand management policy agendas.

2.2. Following principles of sound research design

Attributing changes in practices (such as in washing machine usage patterns) and actual outcomes (such as electricity use) to a particular behaviour change intervention ideally requires an experimental design which controls for other possible explanatory variables [4]. A long time period for the study also helps to assess the duration of any observed changes. This implies a large sample size – IDEAL is aiming to install the sensor system in around 300-400 homes. A large sample size is also important for the other key aims above, helping to test the robustness of the sensor system in a variety of home types, and build a large dataset for machine learning and quantitative social science investigations.

At the same time, aims 1 to 3 above also require small-scale development and testing. The sensor system has gone through a period of lab testing and then testing in a small number of homes to refine it and identify any issues with its functioning. Machine learning requires training datasets from a smaller number of homes with a much higher number of sensors installed (e.g. [5]). Meanwhile, creating new feedback designs requires a combination of several kinds of knowledge: (a) social scientific insights into behaviour or practice change, (b) human-computer interaction (HCI) approaches to understand engagement with digital devices, and (c) user-experience design expertise in turning those insights into material, engaging interventions (in our case, specific user interface designs and feedback content). An initial period of co-design and evaluation with small numbers of participants is therefore important to collect qualitative feedback on items such as the intelligibility, usability and usefulness of feedback designs. Only when there is confidence that a design has at least the potential to have a measurable impact is it appropriate to proceed to an experimental study of its efficacy in changing practices and energy use.

A further element of research design is achieving a sample that has the properties required of it. In our project, the aim is not necessarily to test this across a representative sample of the population, or even of the sociodemographic groups included in the study. Instead we aim to involve household types whose aggregate energy savings, if the system were to be rolled out nationally in future, would be most likely to represent a substantial reduction in carbon emissions. The sample should therefore include households which in the wider population (i.e. UK households) represent a substantial proportion of total carbon emissions (or more accurately, inasmuch as it can be identified, a substantial proportion of the potential total reduction in emissions arising from feedback). It should also include sufficient households within each sub-group, defined along dimensions considered to be important moderators of the effect of our feedback, so that we can test how effective are our feedback interventions within each of these sub-groups.

2.3. Grounding in existing literature and theory

Careful selection of variables to measure, and appropriate instruments with which to measure them, is also important to achieve the research aims. Aims 3 and 4 in particular require variables to be collected which enable us to measure and identify not just if changes in outcome variables arise (such as changes in energy use), but also which factors influence that, including which factors influence the effectiveness of feedback interventions. Grounding the data collection strategy in existing theory and literature helps provide a rationale for which variables to measure. The whole range of constructs considered important for understanding a phenomenon (in this case energy using practices) should ideally be measured, and should also inform the intervention design [6]. In IDEAL, we draw on practices theory and behavioural psychology to inform both our feedback design and evaluation methods, which allows us to explore the effectiveness of these potentially opposing disciplinary perspectives in understanding energy using practices and the effects upon them of our feedback.

3. RESEARCH DESIGN: ADAPTING PRINCIPLES TO THE PRACTICALITIES OF REAL-WORLD RESEARCH

We describe our research design below, taking into account the design requirements above, and discussing the challenges of implementing idealised methods ‘in the wild’, and our approach to addressing some of the compromises between sometimes conflicting theoretical and practical considerations that are also likely to arise in other similar studies.

3.1. Study design

The size and duration of the IDEAL project enables us to develop a novel sensor system, new machine learning methods, data collection methods, and feedback features. However, the development time involved in these stages means they cannot be developed in a serial way; that is, there is not sufficient time to develop and test all these components first at the small scale, then release them all at once to an experimental group. Instead the feature development and experimental testing phases run in parallel, with features developed and small-scale tested, then released to the experimental group, in distinct modules. As much as possible, feedback feature development is being timed to take into consideration factors such as

seasonality, logical sequencing of features, and readiness of machine learning disaggregation tools to provide the disaggregated data upon which they may rely (for example, features related to heating practices and the resultant impact on room temperatures and energy use are being timed for release to the main study group during the heating season, but also require machine learning algorithms to be ready which reliably infer patterns of central heating use).
We therefore subdivided recruited households into a qualitative group of around 40 homes for small scale studies, and a larger experimental sample (~250-350 homes) split into control and treatment groups, with each home intended to be retained from installation day until the end of the project. As needed for the development of individual modules, we also included other short term participants for involvement in focus groups (for the co-development of feedback design) and system installations (for early testing of equipment and installation processes).

3.1.1. Qualitative study group

The main qualitative study group serves a dual purpose. Firstly, for machine learning, these homes provide a training dataset, as they have a much larger number of sensors to directly measure and report usage of individual appliances, including all electrical appliances that typically use most of a home's electricity (via plug level monitors or monitoring of individual circuits of the home), and all hot water outlets, radiators and other gas using-appliances such as ovens and hobs (via temperature sensors).
Secondly, these homes are also being offered feedback features much earlier, enabled because the disaggregated data which the features require are directly measured rather than being inferred by machine learning algorithms (which take time to develop). For feedback feature development, we follow a general approach of an initial co-development phase, in which we draw on existing literature in the field, HCI expertise, and focus groups to co-design the initial mock-ups of features, and then develop them. We then release them to this qualitative study group to collect feedback from them to evaluate the features' intelligibility, usability and usefulness. After a final period of refinements, the features are released to the experimental treatment group at a suitable time.

3.1.2. Experimental study group

The experimental study group also serves various purposes. Firstly it provides a validation dataset for machine learning models. Secondly it is used for testing feedback feature effects. Thirdly is for the sociological study of which other, non-feedback, factors shape a household's patterns of energy using practices, and how these practices shape energy use and other outcomes.
The "gold standard" for experimental research in medicine is the double-blind study, where neither participant nor researcher knows who is receiving the treatment or is in the control. In IDEAL, the control and treatment groups are being recruited with the same offer and receive the same interactions as each other. This is ensured not least because we do not ourselves allocate households to control and treatment groups until after they have been recruited and had the feedback system installed in their homes (see below). Post-installation, households will continue to receive the same interactions with the research team, consisting of a mix of

occasional newsletters, emails relating to new feature releases, access to a phone and email support service, and ad hoc contact as required to address specific system problems that might arise. Of course, the treatment group will receive different updates to the feedback system to the control group, but the recruitment material is framed to be vague about what features will be released and when. We cannot fully control all possible channels by which a household might discover which group they are in however: they might, for example, happen to know other participants, or read a publication of ours describing features which they know they have not received. However, 'the effects of repeated social interaction with the research team are controlled for, allowing us to rule out "Hawthorne" effects of social interaction *per se*' [1]. In real world social experiments, study groups are not strictly controlled – there will be differences in characteristics between the treatment and control groups because of natural variations in the population, so we in reality have a 'quasi-experimental' design [1, pp. 23-24]. Even with large sample sizes these differences effectively reduce the standardised effect size of our intervention, reducing the study's power to detect the interventions' effects, by increasing the standard deviation in the change in the dependent variable. This potentially occurs through these variables directly influencing the dependent variable, or by mediating the effect of the intervention on the dependent variable. Section 3.2 below discusses how we draw on existing relevant theory to guide our understanding of which variables are likely to be important in this respect. A second issue is how to address variance in these variables in the study. There are three options available to control for the variation in these variables:

1. Exclude certain values or value ranges for a given variable.
2. Quota allocate on the variable: This helps ensure a desired distribution of values on these variables (and combinations of them), and that approximately equal numbers of each type of home are allocated to each of the control and treatment groups.
3. Control for these variables in analyses: Measure the variables, not to constrain recruitment or quota allocate, but instead to include them in analyses to control for their effects (see 3.3 below on data analysis) [7, pp. 23-25].

In terms of excluding certain households (1 above), recruitment to the experimental study group is subject to various technical constraints which aim to:

- reduce the difficulty of developing ML models and tailored feedback, e.g. heating system type is constrained to be gas combi boilers.
- increase the quality and duration of the data collected, including factors which might impact on the functioning/robustness of the sensor and feedback system, e.g. excluding those with a self-reported intention to move home, or renting, no broadband, poor WiFi propagation or pre-pay electricity meters.

Constraining recruitment further by introducing other non-technical, theoretically informed, constraints was considered to be problematic for IDEAL, as each additional constraint increases the difficulty in recruitment sufficient numbers of households to the study.

Quota allocation (2 above) is used to control variance along three other dimensions: household income, household composition, and recruitment date. Household income and composition are expected to be significant intervening variables in the impact of personalised behavioural feedback on energy demand. An unavoidably long recruitment period (an estimated 6 to 8 months to recruit the full sample of homes) means that different homes are in the study for different periods before feedback features are released, which could also significantly affect responses to them. Quotas are therefore being allocated along the following dimensions:

- Study group: equal numbers within both control and experimental group (this maximizes study power for any given sample size).
- Recruitment period: equal numbers of homes from each recruitment period should be allocated to the two groups – in IDEAL we do this in two-monthly blocks.
- Household income x household composition - 2x2 cells, as shown in table 1 below, with equal numbers from each cell to each study group for each recruitment period.

The cells are designed to allow us to evaluate how the feedback loop's effectiveness is moderated by the independent variables of household income (above and below UK median income, equivalised per occupant) and household composition (one adult occupant only; more than one occupant) – a 2x2 within-feedback-category design as shown below, leading to 1/8th of the full sample in each sub-cell (38 per sub-cell, assuming a total sample size of 304 homes of usable data). Including single occupancy households as a specific category was decided because it provides a substantial proportion of the sample where we can be fairly certain that inferred practices have been performed by a specific individual, and hence linked to their survey responses and feedback use patterns.

Table 1: Within-feedback-category quota allocation of households

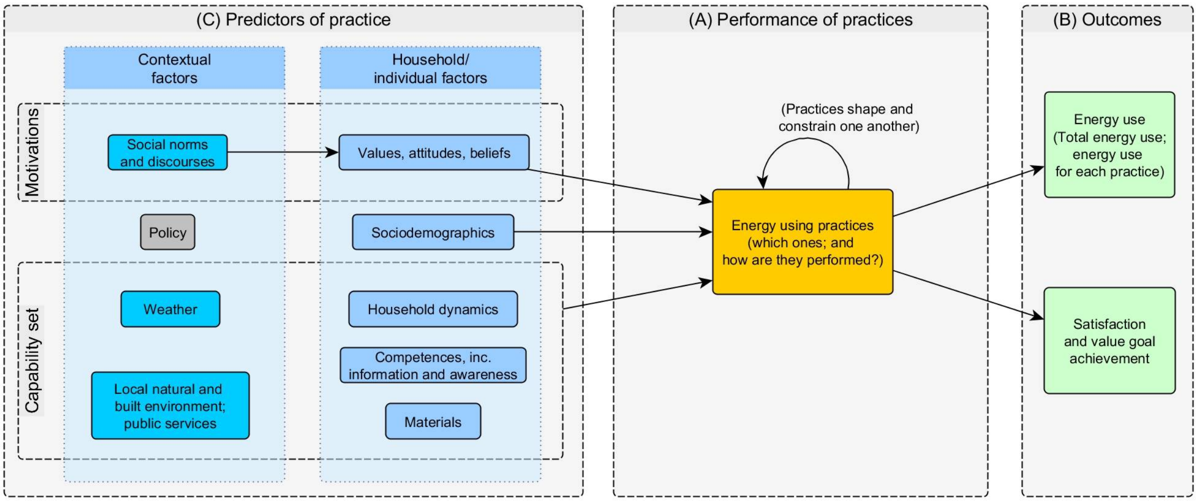
Variables	**Single occupancy only** (1 adult, no dependent children)	**Families** (1+ adults, 1+ dependent children)	**N**
Lower £	38	38	76
Higher £	38	38	76
N	76	76	152

3.2. Data collection

Research in practice theory and behavioural psychology provides insight into variables to measure: energy feedback has been found to have effects, for example, not just (or even) on domestic energy use, but also on how energy-using practices are performed and change over time, on levels of energy literacy, and on energy-related attitudes and motivations. Further variables (such as measures of the socio-demographic characteristics of the occupants, and available household appliances) help to understand performances of practices, and why an intervention did (or did not) have measurable impacts.

A detailed description of the data collection methodology and underlying theories upon which it draws is described in [10]. Figure 1 below summarises our conceptual model, incorporating constructs from the two sets of theory. A range of factors (including our own feedback system, as well as participants' characteristics and material properties of the home, from building fabric to appliances available), at multiple levels, predict practice performance (e.g. frequency, duration, timing, and technology used for washing). Practice performance meanwhile influences a range of measured outcomes, including energy use, associated financial cost and carbon emissions, objective health outcomes, and self-reported comfort, cleanliness, convenience and satisfaction with the practice.

Figure 1: Conceptual model of how domestic energy using practices shape outcomes, and are influenced by different factors



The data collection strategy aims to collect measures for all these constructs. Across all the variables selected for the evaluation, the most appropriate measurement instruments vary. This can lead to a challenging data collection strategy. In particular, measuring change in practice performance is notably difficult using conventional social science methods – time use diaries have a high participant burden, whilst surveys produce only estimated averages of actual practices. One aspect of the IDEAL evaluation involves considering participants as agents who provide data passively (e.g. via sensors and through their patterns of usage of the feedback system), as well as actively, such as through survey completion, interviews, and by engaging in the co-creation process for developing the feedback user interface and content. A key aim of the IDEAL project is to contribute methodologically in this areas, by drawing on machine learning methods to provide quantitative measures of the performance of practices that can be inferred from the traces they leave in the sensor data collected [9] – that is, in the recorded patterns of usage of electricity and gas, and the temperature and humidity levels of rooms in the home. Whilst inferring patterns of appliance use is an area of ongoing work in the machine learning literature (e.g. [5],[10]), incorporating such methods into social sciences models of the home that focus on practice performance, and combining them with data from more conventional social science data collection tools (notably surveys) contributes to a parallel literature on the quantitative measurement of practices (e.g. [8],[9],[11]).
In IDEAL we therefore combined the following data collection methods, in a manner designed to innovate in the quantitative measurement of practice performance and impacts, and collect repeat measures of all key constructs in a way that minimised participant burden and the associated risk of dropout:

- Surveys, collected during installation and six monthly, focusing on household and individual factors listed in (C) in figure 1.

- Sensor data and machine learning inferences of practice performance and impacts: at a resolution of seconds, collecting most of (B) and (A) in Figure 1.
- Questions asked via the feedback system to all participants: Ad hoc, mostly on aspects of practice performance (A) and subjective perceptions of them (B) that cannot be measured or inferred from sensor data (e.g. how often clothes are worn before washing; satisfaction with the thermal comfort of the home).
- Secondary environmental and spatial data: contextual factors (C), including external weather conditions, proximity to greenspace and community spaces, and other factors which might influence time spent in the home).

The resultant dataset includes longitudinal data on: practice performance at high temporal resolution (seconds); predictors of practice performance at individual, household and local scales; and the disaggregated and aggregated energy impacts of those practices.

3.3. Data analysis

Data analysis methods will vary depending on which of the specific research questions are being considered. As discussed earlier, for evaluating the feedback, it is of theoretical interest to consider not just the effect on final energy use, but on a range of other variables too: effects on energy use for particular practices (such as personal washing, heating practices, doing the laundry), on how those practices are performed (timing, frequency, duration, technology used, etc.), on other outcomes of the practices (e.g. temperature and humidity in the home), as well as participants' attitudes towards and knowledge about their energy using practices. In this instance, an experimental method aims to detect if the dependent variable is statistically significantly higher or lower over a given period in the households in the treatment group compared to the control group, either overall or for one of the sub-groups of homes. Given the probably high levels of variance in multiple predictor variables as discussed above, controlling for these is also important – ANCOVA methods, for example, allow these covariates to be controlled for (inasmuch as they have been measured). Each household's baseline value for the *dependent* variable in question can also be included as a control variable in an ANCOVA test as this effectively controls for the direct effects of unmeasured variables too [12]. As households receive feedback that is tailored in both type and quantity, and may have differing levels of contact with the research team, there is effectively within-group variation in 'treatment' too, unlike a drug trial [7]. It can therefore be appropriate to control for this in analyses by including adequate proxy variables of both the types of feedback and the quantity of each type that are delivered, on the one hand, and "consumed" by the household, on the other, as well as measures of research interactions. Care also needs to be taken to address the issue of confounding effects arising from the use of measured variables to target feedback: the intention is that feedback will be tailored to households based on certain of their characteristics, which we might also wish to use as control variables in analyses.

The confidence intervals applied are also relevant in experimental research. The statistical power (the probability of correctly detecting a difference between the control and experimental groups due to an intervention) of evaluations of the study are determined by:

- The sample size
- The alpha level – the threshold sensitivity for Type I error (i.e. probability of failing to identify a genuine effect of the feedback of a given effect size) that is chosen,
- The statistical test, and
- The standardised effect size [12].

The standardised effect size (ES) is given by the formula:

$$ES = \frac{\mu_e - \mu_c}{\sigma}$$

Where μ_e is the mean change in the dependent variable before and after treatment for the experimental group; μ_c is the mean change for the control group; σ is the standard deviation of the change in the dependent variable for the whole sample.

Decisions on alpha and beta levels (sensitivity to type I and type II errors) can be made at the analysis stage, as can choice of statistical test. Whilst confidence intervals (CIs) of 95% or higher are considered appropriate to report in regression analyses, in experimental studies much lower CIs in small effect sizes are often considered worthwhile reporting. Inasmuch as the ultimate aim is to inform future research directions, policy and practice, then low CIs are not necessarily problematic – policy decisions, for example, must often be made under conditions of high uncertainty, in which case low confidence evidence is preferable to none at all.

Finally, the full dataset can be analysed using non-experimental analytical methods to address other research aims not related to the effects of feedback. This includes other factors such as whether distinct and stable categories of variant of specific practices are detectable (for example, using clustering methods to define variants of laundry washing), which factors predict energy use or patterns of practice performance (e.g. using regression or multi-level modelling techniques) and how changes in variables interact over time, such as stability or change in a households' practices (e.g. using Granger causality tests).

4. DISCUSSION

The research design described above was developed to be sufficient to address our research aims, within the context of the complexities of undertaking research 'in the wild', that is, outside the lab, in only weakly controlled conditions. Although various risks were identified and addressed by this approach, there are still further risks that require flexibility in the research design.

One of the principle risks is that there are multiple strong dependencies between the different stages of this modular design approach. Developing a new sensor system to the point where it is reliable and robust enough to install in hundreds of homes is a significant research and design challenge, and delays here impact on the availability of participants and of data for machine learning and feedback. Similarly, uncertainty about the extent to which the sensor data being collected can be reliably disaggregated creates issues for selecting feedback designs to focus on, as they rely on disaggregated data being produced.

This implies the need for flexibility in methods and even research aims, to respond to changing opportunities and events. One early example in the IDEAL project is that it became clear that technical requirements for reliable broadband and mains electricity precluded households with pre-pay electricity meters from participating, so that we could not study this group despite research interest in investigating the potential of the system to support those in fuel poverty (many of which do have pre-payment meters).

A further risk is that various aspects of the project remained unknown until tested at scale. For example, although we draw on the experiences of previous projects and project partners with expertise in participant recruitment, we had no prior data on the likely rates of dropout of participants from the project, nor of the robustness of the data collection system in a wide diversity of home settings over extended periods, and hence there is high uncertainty about the quality of data that can be collected and the capacity to test feedback designs in the manner planned.

Further flexibility is required because this is a field in which there is rapid development in both research and commercial spheres. In our project, an early intention to have a third study group focused on social comparison feedback was removed because we felt sufficient other research had been conducted into the topic that it no longer justified a strong focus in IDEAL.

The modular structure of the project is in part intended to address some of these issues – although there is a working plan for the types of feedback to be developed and tested, as features are being developed and released in modules there is flexibility to change what those modules are in response to issues arising.

5. CONCLUSIONS

We conclude by highlighting how our experiences and approach can help inform the design and evaluation of future experimental research studies into the efficacy of behaviour change interventions, particularly energy related ones. As a starting point, it is important to clearly define the intended research aims and contributions, which may, as problem-focussed research is frequently multi-disciplinary, span various academic disciplines. Equally important is to soundly ground the research design in the principles of research methodology, and to ensure that the data collection and analysis methods robustly drawing on appropriate theories to inform the constructs which should be measured, the appropriate instruments to measure them with, and analytical tools to maximise study power and generate valid conclusions.

Unless a project is testing a system that has already been fully developed and small scale tested, then a develop and small-scale testing phase is also important to avoid risks of releasing technology to the treatment group that fails because of basic issues with reliability, intelligibility, usability and usefulness. Such smaller scale studies also provide further potential for useful project outputs. For projects of long duration, a modular approach to designing features also helps the project to adapt to changing circumstances within the project and in the wider research, commercial and policy environment in which it is involved – such an approach retains flexibility in the project to respond to delays and changes in knowledge and state of the art in the field of study.

Ultimately, following good research design principles helps ensure that projects produce sound scientific knowledge and contribute to their fields of study even in the light of changing conditions, and ultimately helps improve the quality of the research base upon which policy decisions can be made and technology designed.

REFERENCES

[1] N.H. Goddard, J.D. Moore, C.A. Sutton, J. Webb and H. Lovell, "Machine Learning and Multimedia Content Generation for Energy Demand Reduction". *Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–5, (2012).

[2] T. Erickson, M. Li, Y. Kim, A. Deshpande, S. Sahu, T. Chao, P. Sukaviriya and M. Naphade. "The Dubuque Electricity Portal: Evaluation of a City-Scale Residential Electricity Consumption Feedback System" *CHI*. pp. 1203-1212, (2013).

[3] T. Harries, R. Rettie, and M. Studley. "Using Digital Technologies to Test the Social Norms Approach to Reducing Electricity Consumption", *Proc. HCI* (2012).

[4] J. Froehlich, L. Findlater, and J. Landay, "The Design of Eco-Feedback Technology". *Proc. 28th Int. Conf. Human Fact. in Comp. Sys.*, pp. 1999–2008. 2010

[5] M. Zhong, N.H. Goddard, and C.A. Sutton, "Interleaved Factorial Non-Homogeneous Hidden Markov Models for Energy Disaggregation", *NIPS*. Lake Tahoe, (2013)

[6] E.B. Hekler, P. Klasnja, J.E. Froehlich, and M.P. Buman, "Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research", *Proc. CHI*. New York, (2013).

[7] F. Bechhofer and L. Paterson, *Principles of Research Design in the Social Sciences*. London: Routledge.

[8] M. Pullinger, "The Quantitative Measurement of Social Practices in the Home, Their Influences, and Their Impacts: A Mixed Methods Approach Using Innovative Technologies." *Working Paper*, University of Edinburgh, (2016).

[9] W. Medd and E. Shove, *The Sociology of Water Use*. London: UK Water Industry Research, (2006)

[10] J. Liao, L. Stankovic, and V. Stankovic, "Detecting Household Activity Patterns from Smart Meter Data", *IEEE*, pp. 71–78. (2014).

[11] M. Pullinger, B. Anderson, A.L. Browne and W. Medd, "New Directions in Understanding Household Water Demand: A Practices Perspective". *J. Wat. Supp. Res. Tech.- AQUA* 62(8): pp. 496–506, (2013).

[12] M.W. Lipsey, "Design Sensitivity: Statistical Power for Applied Experimental Research". L. Bickman and D.J. Rog eds. Handbook of Applied Social Research Methods, pp. 39–68. Thousand Oaks: Sage Publications, (1998).