Evaluating Energy Culture: Identifying and validating measures for behaviour-based energy interventions

Rebecca Ford, SEE Change Institute, Los Angeles, CA
Beth Karlin, SEE Change Institute, Los Angeles, CA
Cynthia Frantz, Oberlin College, Oberlin, Ohio

Abstract

Energy behaviour is embedded within the physical and social contexts of daily life; the interplay between behaviour and its contextual influences can be thought of as an "energy culture". Behaviour-based energy interventions aim to impact demand through influencing some aspect of energy culture - what people have, think, and/or do. Understanding how a program does (or doesn't) work requires an understanding of changes in these elements of energy culture. This paper presents and tests a set of instruments that evaluate household energy culture before and after an intervention. Findings indicate that the proposed instruments are consistent with the conceptual variables they were designed to measure and have strong predictive validity. This work highlights the potential for reliable and valid methods to complement traditional measures of program effectiveness, providing deeper learning into how interventions lead to savings. Such insights can support program improvement, and ultimately increase the impact of behavioural interventions.

Introduction

Energy systems in many countries across the globe are changing in response to the "trilemma" of meeting energy security, energy equity, and environmental sustainability goals (World Energy Council, 2015). Alongside the $53 trillion of investment in energy infrastructure needed by 2035 to meet a 2 degree climate change goal, behavioural strategies offer substantial opportunity to cost effectively support emission reduction mandates and may aid integration of time variable renewable energy generation (Allcott & Rogers, 2014; Dietz et al., 2009; Dupont et al., 2014). In light of this, a substantial number of programs and interventions have been implemented to explore the impact of behaviour-based energy programs, yet results vary and there are still many unknowns (Karlin et al., 2015a, Karlin et al., 2015c; Lutzenhiser et al., 2009).

In order to understand how interventions work best and identify potential program improvements, we need to evaluate these programs to explore how interventions are leading to behaviour changes and in what contexts (Karlin et al., 2015b). In this paper, we present and test a set of instruments designed to support traditional evaluation metrics (e.g., energy and/or gas usage) by providing additional insights into the pathway for achieving reductions in demand and enabling comparisons across programs. The work presented here is the subset of a larger project, conducted in conjunction with the International Energy Agency Demand Response Programme (IEA-DSM) Task 24 on Behaviour Change (see Karlin et al., 2015a), and focuses specifically on household energy behaviour and change.

Understanding Behaviour

Understanding energy behaviour has become increasingly important for both government and industry, so that they are able to develop and implement more effective policies and programs utilising strategies that target end-user behaviour to reduce energy and peak demand (Chatterton, 2011; Todd et al., 2012). Exploring why people behave in certain ways and how change is facilitated requires a consideration of people, their actions, and the social and physical context in which behaviour occurs. Scholars have studied behaviour and behaviour change through a variety of different disciplinary lenses across both the physical and social sciences (e.g., Chatterton, 2011; Froehlich et al., 2010; Geels, 2010; Lutzenhiser, 1993; Wilson & Dowlatabadi, 2007). Each of these disciplinary perspectives provides a different view of behaviour and change, though none on their own explore the full range of behavioural predictors.

Some models go beyond one aspect of behaviour and explore multiple drivers of behaviour. For example, the ABC model (Guagnano et al., 1995) explores the relationships between external conditions, attitudes and behaviours, and attempts to bridge distinct approaches taken by social psychologists and behavioural economists. However, these models fail to consider the heterogeneous characteristics of energy behaviours as they span different contexts and scales (Keirstead, 2006; Wilson & Dowlatabadi, 2007). Consequently they are rarely used in practice, despite the need for "energy policies and programs ... to replace outmoded assumptions about what drives human behaviour" by taking an approach that integrates "insights from the behavioural and social sciences with those from engineering and economics" (Dietz et al., 2013: pp 78). Although this approach has only occasionally been implemented, it is urgently needed to be able to tap the potential for individuals and households to reduce carbon-based energy consumption (Dietz et al., 2013).

The Energy Cultures Framework (Stephenson et al., 2010; 2015) was designed to offer an integrated approach across different disciplinary perspectives of energy behaviour, including the acquisition, use and divestment of energy-related technologies that impact upon energy consumption. It provides a structure to explore how people and their energy behaviours are embedded within the physical and social contexts of daily life, and both shape - and are shaped by - external influences such as technological developments, regulatory environments, policy settings, market forces, cultural meanings, social practices, and preferences (see Figure 1). This interplay between actors, energy behaviour and the wider contextual influences can be thought of as an "energy culture".

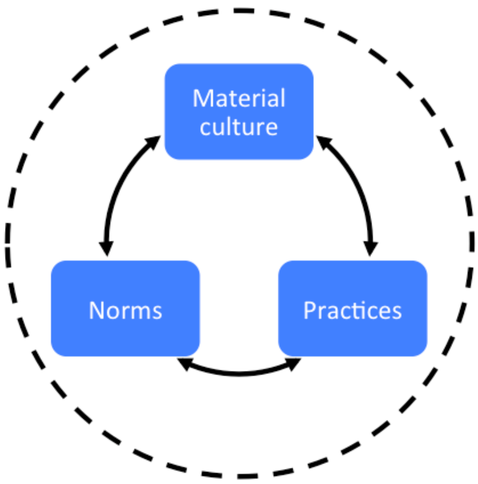


Figure 1. The Energy Cultures Framework

Understanding a household's energy culture requires a consideration of the things they have (e.g. heating devices), what they do (e.g. setting the thermostat), and how they think (e.g. expectations around comfort), as well as the interrelationships between these elements, and the interactions between each element and external influences. Behaviour-based energy programs are able to impact energy use through influencing one or more aspects of a household's energy culture. These programs can yield significant energy savings, but to understand why or how the program was successful requires not only measuring the kWh reduction in demand, but also exploring the corresponding shifts in energy culture.

Developing the Measurement Framework

In creating a set of instruments to measure shifts in energy culture, it is important that the tool captures constructs of interest in accurate and reliable ways (Karlin et al., 2015b). To do this requires developing and validating measures in a precise manner, as illustrated in Figure 2.

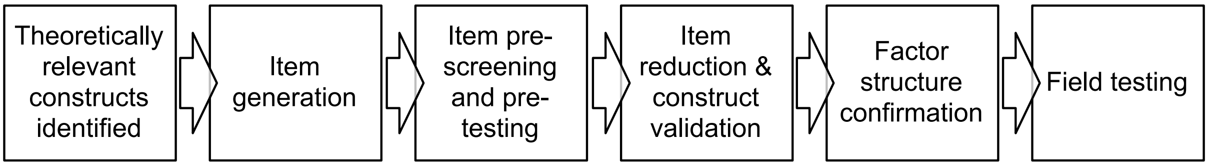


Figure 2. Process to develop measurement instruments

The first step involves conducting a literature review to identify relevant constructs from past theory and empirical research. The measurement framework proposed here is designed to explore shifts in each aspect of a household's energy culture following a behavioural intervention. Questions were developed within each of the four themes: context; material culture; practices; and norms. A series of theoretically and empirically relevant constructs that sit within each of these four themes were identified in prior work (Karlin et al., 2015a; Karlin et al., 2015b). Table 1 details the key constructs within each of the four evaluation themes and the reason for their inclusion[1]. Our literature review focused primarily on household energy use, however we anticipate additional constructs of interest beyond those detailed in Table 1. Thus while we take a broad interdisciplinary perspective in framing the toolkit, the scales tested in this paper may be only a subset of the overall framework. Ultimately, we view the toolkit development as an on-going process with the addition of further constructs and instruments over time (following the approach presented here to developing and validating measures) and from which researchers and evaluators may borrow to craft an intervention appropriate evaluation measurement that incorporates rigorously validated instruments.

Having identified constructs, the second step is item generation. This involves compiling a large pool of potential items based on reviewing questions used in previous research, and generating new items for constructs inadequately represented. In developing specific questions to evaluate each construct, we drew from prior work using the following guiding principles:

(1) Items that had been developed and psychometrically validated by others were used verbatim;
(2) Items that had been developed by others but not psychometrically validated were reviewed by the research team alongside similar items and modified as appropriate;
(3) Items where no previous questions had been published were developed by the research team with reference to other related items and context of testing.

This initial work was developed for implementation in California and questions were designed with this context in mind. It is anticipated that broader use of this toolkit may require some instruments to be revised. For example, the material culture and practices items should reflect those appropriate to the region within which the intervention occurs. Further response items in these sections may also need to be added depending on the specific intervention being evaluated; while the current work has been developed for a broad evaluation of residential energy demand in Californian homes, the addition of mobility or small business focussed practice and material culture items would enable the toolkit to be used in additional contexts beyond residential demand.

1 Full references available from authors on request

Table 1. Key constructs identified for testing

Theme	Construct	Description
Context	Physical context	Behaviour is often constrained by physical and structural realities of a dwelling and its technologies
	Demographics	This helps inform for whom an intervention works best for, and can identify cultural constraints on change
Material Culture	Appliance ownership	Understanding changes in appliance ownership over time can help explain energy use changes
Practices	Recurring Behaviours	Taking a baseline measure of habitual behaviours that people engage in allows a precise measure of change
	One-time Behaviours	Taking a baseline measure of one-time behaviours that people engage in allows a precise measure of change
Norms	Drivers of use	Different factors govern the reasons for and use of energy
	Motivation to Save	Different factors govern the reasons people try to save energy
	Energy Literacy	Awareness of the larger context in which behaviour occurs can promote better decision making
	Energy Knowledge	Knowledge specific to a particular domain is a predictor of behaviour in that domain
	Concern and Connection	People are more likely to take action if they feel connected to the environment or concerned about it
	Personal Norms	Personal norms are likely to motivate behaviour change when it is perceived that the environment is threatened
	Social Norms	Believing that other people are engaging in/approve of a behaviour is a strong predictors of that behaviour
	Performance Efficacy	Individuals must believe they have the ability to perform a behaviour or they will not attempt to do so
	Response Efficacy	Individuals must believe that the behaviour will have its intended effect or they will not attempt to engage

In the third step items are pre-screened and prepared for testing through review by researchers and practitioners, then reviewed by lay audience for clarity of wording and appropriate answer choices. The resultant toolkit from this stage of testing contained 158 items across the themes of context (8 items about physical context, 11 demographic), material culture (17 items about appliance ownership), practices (14 items about recurring behaviours, 29 about one-time behaviours), and norms (10 items about drivers of energy use, 16 items on motivations to save energy[2], 5 on energy literacy, 7 about knowledge, 10 related to connection and concern, 8 on personal norms, 6 on social norms, 9 about performance efficacy, and 8 to gauge response efficacy). The initial item set was designed with significant redundancy so multiple questions per construct could be tested and only the strongest retained for field implementation.

Steps 4 and 5 describe the testing that is done to ensure that: items proposed reliably and accurately measure the constructs they are designed to, are useful in predicting behaviour, and hold true across different populations.

---

2 Participants who said that they do not try to save energy in their household were not asked the motivation to save question. The items on motivations to save were presented to a randomized subset of the study sample as part of an exploratory analysis (other participants were presented with a variation of the question, such as in an open-ended format, to examine if our scale is inclusive and tapping key motives).

The final step in validation is field testing, in which the measures developed are used in field setting to confirm that they are sensitive enough to detect changes in behaviour, and to remove any redundant constructs that have no explanatory utility.

Before sharing the instruments for use, it is thus important that they are tested and validated to: (1) ensure that all respondents interpret the language of the questions in the same way, (2) to ensure that this interpretation matches the intended meaning of the researcher; and (3) to create a final set of questions that are useful for researchers to implement. This work presents testing on item reduction and construct validation for a set of constructs designed to measure energy culture.

Testing and Validation

Participants were recruited via Amazon's Mechanical Turk (see Mason & Suri, 2012) and were paid $1.25 for their time. Of the 348 completed responses, most participants (69.5%) lived in houses; 73.8% of these owned their home, while 22.9% rented. Households had an average of 2.9 members (SD = 1.36). Gender was evenly split, with 49.7% female, 49.7% male (0.6% did not answer the question). The majority of participants (84.8%) identified themselves as White.

Participants completed all potential items identified from the literature review designed to measure the constructs described in Table 1. In addition, they completed several scales to establish convergent and divergent validity. This was done to ensure that the newly developed items were accurately measuring the constructs they had been designed to measure by demonstrating that they: (1) correlated with existing (already previously validated) items that measure related constructs (e.g. New Ecological Paradigm), and (2) do not correlate with existing (already previously validated) items that measure theoretically distinct constructs (i.e., Big Five Personality Inventory).

Participants completed the Connectedness to Nature Scale (CNS, Mayer & Frantz, 2004) and the New Ecological Paradigm (NEP, Dunlap et al., 2000); two widely used measures of how people relate to and think about the environment. Both of these scales were predicted to correlate positively with energy behaviours already taken, behavioural intentions, and concern and connection, and personal norms. Participants also completed the Environmental Attitudes Scale (EAS; Ebenbach, 1999), which includes a subscale that measures internal motivation to be pro-environmental (EAS-I), as well as a subscale that measure external motivation (EAS-E). The latter subscale served as a domain-specific measure of socially desirable responding. Finally, participants responded to a short version of the Big Five Personality Inventory (NEO-Brief, Thompson 2008), which measures extraversion, neuroticism, openness, agreeableness, conscientiousness. While we predicted modest correlations between particular toolkit subscales and personality traits, we primarily included the Big Five to establish divergent validity.

Our primary analysis goal was to identify a concise subset of items that reliably, cleanly and validly measured each construct of interest (i.e. a scale). We used reliability analysis and exploratory factor analysis to reduce the overall number of items within each scale. This created 16 distinct scales, each of which is asked as a single survey question containing between 1 and 4 items that have high reliability and clean factor loadings. We then examined convergent and divergent validity through correlations with established scales. Finally we tested predictive validity through a series of regression equations that tested for predicted relationships between the variables based on well-established theories of behaviour.

Findings

First, all items designed to measure a single latent construct (e.g. all motivation to save energy) were subjected to reliability analyses to ensure that each individual item correlated with other items designed to measure the same construct. To test for reliability, a Cronbach's alpha was computed for each set of items (and the alpha that would result if each individual item was deleted). Reliabilities ranged from acceptable (0.68) to excellent (0.92). No single item significantly decreased the alpha of its scale, so no items were eliminated on this basis.

Next, items designed to measure drivers of use, motivation to save, energy literacy, connection and concern, personal norms, social norms, performance efficacy, and response efficacy were subjected to exploratory factor analysis (using Principal Components Analysis with a non-orthogonal rotation and factor extraction based on eigenvalues over 1). A primarily goal was to reduce the overall number of items measuring each construct. These analyses also allowed us to test whether the items designed to assess each construct cohered along a single dimension (i.e. whether they actually did measure one single construct), or whether the items actually measured related but slightly different constructs. When constructs were related to one another (i.e. drivers of use; motivations to save; performance efficacy and response efficacy; personal norms and social norms), we ran the analysis combining the related constructs to ensure that the items selected distinguished clearly between the two concepts. Factor loadings were used to identify and validate subscales within each construct; they were also used to identify the items that best represented each factor, and to eliminate items that did not load well. When items loaded equally well, an effort to include reverse-scored items was made.

Factor analyses revealed that the vast majority of our original items loaded in ways that were consistent with the conceptual variables being measured. The analysis of drivers of use yielded three main factors: pro-social motivation, self-comfort, and financial cost. Motivations to save yielded five factors: learning/challenge, self-comfort, social norms, cost, and environmental impact. Connection and concern items yielded three factors: connection, concern, and systems thinking. The personal and social norm items loaded onto two factors, as did performance and response efficacy items. For each construct, we identified two to four items that loaded strongly on the main factor and did not load on other factors. In other words, each item selected for use in the condensed scale clearly related to the relevant construct and was independent from similar constructs. The reliabilities of the condensed scales were comparable to the original scales, suggesting that internal reliability was not sacrificed for brevity (see Table 2).

To establish both convergent and divergent validity of the toolkit items we compared the condensed version of each construct to the established scales described above. Prior to analysis, we established a series of a priori predictions based on theory. Correlation analyses revealed that the vast majority of our predictions were supported, suggesting that the condensed scales (see Appendix A for an example of the validation process) do in fact measure the constructs of interest.

We paid particular attention to the relationship between the toolkit constructs and both EAS-E, a previously validated measure of people's tendency to engage in environmentally responsible behaviour because of social pressure from others, and EAS-I, a measure of people's tendency to engage in environmentally responsible behaviour because of intrinsic motivation. The EAS-E did not correlate with self-comfort or cost drivers of use, cost motivations to save energy, energy literacy, or systems thinking. The EAS-E correlated positively and significantly with connection, concern, personal norms, social norms, response efficacy, and pro-social motivation (r's ranging from 0.25 to 0.53). However in many cases, except pro-social driver of use and the four motivations to save energy, the correlation between EAS-I and the variable was much stronger (usually two to three times stronger) than the correlation between EAS-E and the variable. This suggests that internal motivations for pro-environmental behaviour are more predictive of responses to toolkit items than motivation to appear environmental. Including several items from the EAS-E in future iterations of the toolkit would allow researchers to control for the small percentage of variability explained by socially desirable responding.

Table 2. Reliabilities of condensed scales

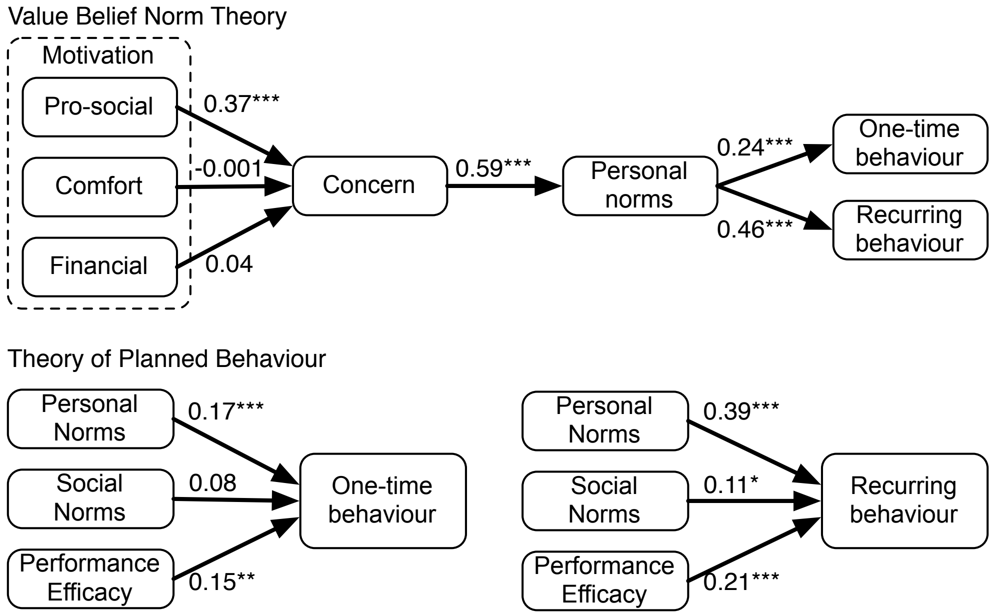
Construct	# items	N	Alpha
Energy Literacy -- Awareness	3	345	0.879
Connection	2	346	0.721
Systems Thinking	2	348	0.751
Concern	2	346	0.713
Personal Norms	3	344	0.818
Social Norms	2	345	0.714
Performance Efficacy	2	346	0.687
Response Efficacy	3	346	0.842
Motivation to save: Learning/challenge	3	75	0.770
Motivation to save: Self-comfort	3	75	0.867
Motivation to save: Social emotions	2	74	0.809
Motivation to save: Pro-social	2	75	0.699
Drivers of use: Pro-social	3	347	0.761
Drivers of use: Self-comfort	3	346	0.695

Drivers of use: Cost and *Motivation to save: Cost* are both single item measures so reliabilities are not computed

The final test of the toolkit items is whether they predict behavioural intentions and behaviour. Future research will need to test the toolkit in a field setting with actual measures of behaviour. However, because the current data set included self-reported behaviours and intentions, we were able to do preliminary tests on the effectiveness of toolkit items to predict behaviour in theoretically expected ways. Three dependent variables were used to conduct these analyses. First, we totalled up the number of one-time energy saving behaviours participants reported having already taken. Second, we averaged together the frequency with which participants reported engaging in recurring energy saving behaviour. We also had a measure of behavioural intention. We totalled the number of one-time behaviours participants said they plan to do.

We used the Value Belief Norm Theory (Stern, 2000) and the Theory of Planned Behaviour (Ajzen, 1985) to guide our predictive validity testing (findings shown in Figure 2). The VBN theory posits that values lead one to become concerned for the valued object (in this case, the environment), which in turn lead the individual to develop personal norms for taking action. These norms then lead to behaviour, in this case energy reducing behaviour. We conducted path analysis with the condensed scales predicting both one time and recurring behaviour. As predicted by the model, pro-social values and cost concerns both predicted concern, which in turn predicted personal norms. Personal norms significantly predicted one-time behaviours already completed and the frequency of engaging in recurring conservation behaviours.

The TPB predicts that social norms, attitudes, and efficacy (specifically performance efficacy) all contribute to forming behavioural intentions, which in turn predict behaviour. Because of the cross-sectional nature of the data set we were unable to test whether intention leads to behaviour. However, we were able to test whether social norms, attitudes, and efficacy predict self reported behaviours. The condensed scales measuring social norms, personal norms, and performance efficacy were tested to see if they successfully predicted recurring conservation behaviours and one-time behaviours. Findings suggest that all three factors significantly predicted engagement in recurring conservation behaviours. All but personal norms significantly predicted one-time behaviours; the effect was in the right direction but not significant. The weaker results for one-time behaviours may be explained by the fact that they are typically more contextually constrained than recurring behaviours. For example, both renters and owners can lower the thermostat (a recurring behaviour), but only owners can make permanent improvements such as installing insulation (a one-time behaviour).



*Note: * p<0.05; ** p<0.01; *** p<0.001*

Figure 2. Predictive Validity Testing

Overall these results suggest that the condensed scales have good predictive validity. The scales predict each other, behaviour, and behavioural intentions in theoretically predicted ways.

Discussion

The goal of this project is to develop and test a set of instruments designed to evaluate household energy culture. Consistent use of such instruments will enable researchers and evaluators to understand how and for whom behavioural interventions are effective by exploring how an intervention leads to shifts in energy culture, and ultimately to shifts in demand. They are designed for use in both pre- and post- evaluation of behaviour-based energy interventions. To ensure that the items developed are reliably interpreted, measure the constructs they are intended to, and predict behaviour intentions, it is important that they are validated before being shared for wider use. This paper presented the first part of this validation.

Findings indicate that the proposed instruments are both consistent with the conceptual variables they were designed to measure and have strong predictive validity. This work highlights the potential for reliable and valid methods to complement traditional measures of program effectiveness and provide deeper learning into how interventions lead to savings. Such insights can support iterative improvements to program design as well as aiding learning across programs, and increasing the impact of behaviour-based interventions across programs.

Next steps leading toward a standardized measurement framework includes field-testing and revision of the instruments developed and validated here as well as additional instruments designed to measure other key identified constructs. While the present paper presented several scales, this is by no means an exhaustive list of variables for inclusion in the evaluation of behaviour-based energy interventions. Previous work by the authors has tested an instrument for user experience (Karlin & Ford, 2013) and additional items that may be worth considering in future include social context (e.g., neighbourhood effects, role of intermediaries), customer satisfaction, and non-energy benefits of program participation.

This work is situated within a broader initiative conducted by the International Energy Agency Demand Side Management Programme (IEA-DSM) Task 24 and we are working in conjunction with several universities, utilities, and program implementers to ensure it meets both scientific and applied standards. It seeks to create a dynamic and adaptable toolkit of measures that are vetted for psychometric/scientific validity as well as practical/field validity so that program evaluators everywhere have a set of standardized measures that they can use and trust to meet their specific evaluation needs. While some questions will be context-specific (e.g., demographics, behaviours), others can be applied universally given adequate translation. As such, broad use of such an instrument can improve and aggregate our overall knowledge across the countless additional studies expected to be conducted in the coming years.

Acknowledgements

The authors wish to acknowledge the on-going support of the International Energy Agency Demand Response Programme (IEA-DSM) Task 24 on Behaviour Change, as well as Southern California Edison and Yinsight for their support with prior and related work. The authors also wish to acknowledge Matthew Ballew for his diligent data analysis and contributions to this work.

References

Allen, D., & Janda, K. (2006). The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: a pilot study. In *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104*, 3003–3037.

Ajzen, I. (1985). *From intentions to actions: A theory of planned behaviour* (pp. 11-39). Springer Berlin Heidelberg.

Chatterton, T. (2011). An introduction to thinking about ‘energy behaviour’: A multi-model approach. Department of Energy and Climate Change, London.

Dietz, T., Gardner, G. T., Gilligan, J., Stern, P. C., & Vandenbergh, M. P. (2009). Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. *Proceedings of the National Academy of Sciences*, *106*, 18452-18456.

Dietz, T., Stern, P. C., & Weber, E. U. (2013). Reducing carbon-based energy consumption through changes in household behavior. *Daedalus, 142*, 78–89. Columbia Business School Research Paper No. 13-54.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: Measuring endorsement of the new ecological paradigm: A revised NEP scale. *Journal of Social Issues, 56*, 425–442.

Dupont, B., De Jonghe, C., Olmos, L., & Belmans, R. (2014). Demand response with locational dynamic pricing to support the integration of renewables. *Energy Policy, 67*, 344-354.

Ebenbach, D. H. (1999). Behind beliefs and controversies: The relationship between motivation and the expression of pro-environmental attitudes. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Froehlich, J., Findlater, L., & Landay, J. (2010). The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1999–2008). Atlanta, GA: ACM.

Geels, F. W. (2010). Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Research Policy, 39*, 495–510.

Guagnano, G. A., Stern, P. C., & Dietz, T. (1995). Influences on attitude-behavior relationships: A natural experiment with curbside recycling. *Environment and Behavior*, *27*(5), 699-718.

Karlin, B., & Ford R. (2013). The Usability Perception Scale (UPscale): A measure for evaluating feedback displays. In: *5th International Conference on Human-Computer Interaction*.

Karlin, B., Ford, R., Wu, A., Nasser, V., & Frantz, C. (2015a). *What Do We Know About What We Know? A Review of Behaviour-Based Energy Efficiency Data Collection Methodology*. Report prepared for the International Energy Agency Demand Side Management Program (IEA-DSM) Task 24-Behaviour Change in DSM.

Karlin, B., Ford, R., & Frantz, C.M., (2015b). Exploring deep savings: A toolkit for assessing behavior-based energy interventions. In *Proceedings of the International Energy Program Evaluation Conference*. Long Beach, CA: IEPEC.

Karlin, B., Zinger, J., & Ford, R. (2015c). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin, 141, 1205*–1227.

Keirstead, J., (2006). Evaluating the applicability of integrated domestic energy consumption frameworks in the UK. *Energy Policy, 34*, 3065–3077.

Lutzenhiser, L. (1993). Social and behavioral aspects of energy use. *Annual Review of Energy and the Environment, 18*, 247–289.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23.

Mayer, F., & Frantz, C. (2004). The connectedness to nature scale: A measure of individuals’ feeling in community with nature. *Journal of Environmental Psychology, 24*, 503–515.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. Personality & Social Psychology Bulletin, 34(7), 913–923.

Southern California Edison (2015). Dimensions of Energy Behavior: Psychometric Testing of Scales for Evaluating Behavioral Interventions in Demand Side Management Programs. Project Report Prepared for the Emerging Technologies Coordinating Council. Project number ET15SCE8010.

Staats, H., Harland, P., & Wilke, H. A. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. Environment & Behavior, 36(3), 341–367.

Stephenson, J., Barton, B., Carrington, G., Doering, A., Ford, R., Hopkins, D., ... & Wooliscroft, B. (2015). The energy cultures framework: Exploring the role of norms, practices and material culture in shaping energy behavior in New Zealand. *Energy Research & Social Science, 7*, 117–123.

Stephenson, J., Barton, B., Carrington, G., Gnoth, D., Lawson, R., & Thorsnes, P. (2010). Energy cultures: A framework for understanding energy behaviours. *Energy Policy, 38*, 6120–6129.

Stern, P. C. (2000). Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues, 56*, 407–424.

Thompson, E. R. (2008). Development and validation of an international English Big-Five mini-markers. *Personality and Individual Differences, 45*, 542–548.

Todd, A., Stuart, E., Schiller, S., & Goldman, C. (2012). Evaluation, measurement and verification (EM&V) of residential behavior-based energy efficiency programs: Issues and recommendations, DOE/EE-0734, State and Local Energy Efficiency Action Network. Washington: US Department of Energy.

Wilson, C., & Dowlatabadi, H. (2007). Models of decision making and residential energy use. *Annual Review of Environment and Resources, 32*, 169–203.

World Energy Council (2015). 2015 Energy Trilemma Index: Benchmarking the sustainability of national energy systems. World Energy Council, London. ISBN: 978-0-946121-42-7.

Appendix A: Illustration of the validation process[3]

Steps 1 and 2: Literature Review and Item Generation
As a result of the literature review, the construct "Drivers of Use" was identified as key to measure. The following items to capture the variety of drivers of energy use were identified from a review of past work. Additional items were added as needed (based on a review of related theory).

Item:	**From:**	**Item:**	**From:**
Environmental impact	Nolan et al. (2008)	Health of household members	New
Cost of the energy bill	Nolan et al. (2008)	Societal benefit	Nolan et al. (2008)
Convenience	New	Keeping use similar to others	Nolan et al. (2008)
Habit	New	Moral obligation	Staats et al. (2004)
Comfort	Allen & Janda (2006)	Guilt	New

Step 3: Pre-screening and pre-testing
Items were prepared for testing through review by colleagues, and then via online testing. Questions were presented to participants (who were asked to provide feedback) and refined as needed[4].

Steps 4 and 5: Item reduction and factor structure confirmation[5]
Factor loadings identified and validated subscales within each construct and eliminated items that did not load well or relate well with others (i.e., below .70). The analysis of drivers of use yielded three main factors: pro-social motivation (3 items), self-comfort (3 items), and financial cost (1 item).

Item:	**Pro-social motivation**	**Self-comfort**	**Financial cost**
Environmental impact	***.710***	-.153	.269
Cost of the energy bill	-.021	.136	***.938***
Convenience	-.106	***.786***	.098
Habit	.100	***.754***	.007
Comfort	-.158	***.765***	-.031
Health of household members	.373	.470	.037
Societal benefit	***.790***	-.016	-.042
Keeping my use similar to others	.611	.148	-.248
Moral obligation	***.828***	-.126	.101
Guilt	.681	.102	-.184

Reliability analysis was conducted to ensure that scale items interrelated well. Items that did not would show an increase in Cronbach's alpha if removed (i.e., better reliability), whereas items that related well with others would show a decrease in alpha if removed (i.e., worse reliability).

Scale and items:	**Alpha of scale:**	**Alpha if item removed:**	**Scale and items:**	**Alpha of scale:**	**Alpha if item removed:**
Pro-social motivation			Self-comfort		
Environmental impact	0.761	0.732	Convenience	0.695	0.593
Societal benefit		0.688	Habit		0.607
Moral obligation		0.617	Comfort		0.608

Next steps: Factor structure confirmation and field-testing

3 Further toolkit items can be provided on request
4 See Southern California Edison (2015) report for further examples of pre-screening
5 Convergent, divergent and predictive validity findings are presented in main body of paper.