Dimensions of Energy Behavior: Psychometric Testing of Scales for Evaluating Behavioral Interventions in Demand Side Management Programs

ET15SCE8010

Prepared by:

Emerging Products
Customer Service
Southern California Edison

December 2015

SOUTHERN CALIFORNIA EDISON®
An EDISON INTERNATIONAL® Company

Acknowledgements

Southern California Edison's Emerging Products (EP) group is responsible for this project. It was developed as part of Southern California Edison's Emerging Technologies Program under internal project number ET15SCE8010. SEE Change Institute and Cynthia Frantz, Ph.D., conducted this work under consulting project manager Yinsight, Inc, with overall guidance and management from Edwin Hornquist. For more information, contact edwin.hornquist@sce.com.

Disclaimer

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

Behavioral interventions in demand side management programs (i.e., those targeting savings through consumer energy use) are based on the idea that people can be encouraged to use less energy if the underlying determinants of their behavior are influenced in some way. Research on such programs suggests potential savings, but results vary and much is still unknown about the mechanisms by which behavior-based interventions result in energy savings.

This is due in part to the way these programs are typically evaluated. Most energy efficiency evaluations use the amount of energy use (measured in kWh) as the dependent variable for measuring effectiveness. Although this is an ideal measure of *whether* energy efficiency interventions work, additional information could add significantly to our understanding about *how* and *for whom* they work. While recent efforts have been made to include additional variables (e.g., attitudes, behaviors, user experience), widespread agreement on variable measurement is lacking and current instruments have not been psychometrically validated to test and improve internal validity. Greater consistency and the use of validated instruments would improve our overall ability to account for variation in treatment effects and identify the underlying causes of intervention failures.

As such, this report presents psychometric testing of a set of scales that can be used to collect self-report data as a part of evaluation of behavioral interventions. It builds from preliminary instruments drafted as part of the International Energy Agency Demand Side Management Program Task 24 on Behavior Change (Karlin & Ford, 2013a) by refining and psychometrically validating the following scaled for use in field studies within California.

1. Norms (e.g., efficacy, social norms)
2. Practices (e.g., onetime, habitual)
3. Material culture (e.g., appliance stock)
4. Context (e.g., demographics, housing),
5. User experience (e.g., ease of use, engagement)

A full survey instrument with these scales can be completed via computer, paper or phone in 10-15 minutes.

Chapter 1 introduces the project and preliminary work. Chapter 2 discusses the theoretical rationale behind the inclusion of each scale and presents methods and results of testing procedures. Chapter 3 presents the validated instrument with suggestions for administration and analysis. Chapter 4 concludes the report with a discussion of next steps and future work.

Broad use of such instruments can improve and aggregate our overall knowledge across studies and contribute to a more robust understanding of behavioral interventions as a resource. Such an understanding is necessary to provide consistent and reliable EM&V calculations across behavioral interventions in demand side management programs that can be compared with supply-side sources of energy and valued in practice as much as they are valued in theory.

1. INTRODUCTION

Utility programs in California and beyond have begun to increasingly target behavior as a means to achieve energy efficiency savings. The Department of Energy (DOE) State and Local Energy Efficiency Action Network (SEE Action, 2012) use the term behavior-based energy efficiency programs and define them as "those that utilize strategies intended to affect consumer energy use behaviors in order to achieve energy and/or peak demand savings" (p. 1). These programs are based on the idea that consumers can be encouraged to use less energy if the underlying determinants of their behavior change in some way.

Traditionally, metrics used to evaluate such programs provide a standard approach for assessing whether or not a program has been effective, but provide little insight into these underlying determinants, which could help understand how and for whom such interventions work. Researchers and evaluators tend to be left to develop their own metrics for assessing such factors, making it hard for utilities to gain additional insights by comparing across programs.

A review of data collection methodology across 85 behavioral intervention studies in demand side management programs published between 2003-2013 (spanning commitment, goal setting, audits, media, feedback, and incentives) was recently completed as part of the International Energy Agency Demand Side Management Program Task 24 on Behavior Change (Karlin et al., 2015). The authors found that, while many studies collected some data in addition to energy use, few collected data across all variables of interest. Among the studies that did collect data about similar variables, the specific metrics used to evaluate them were inconsistent, preventing comparisons between and across studies. The report recommends that future evaluations of behavioral interventions include consistent measures of context, behaviors, attitudes, knowledge, and user experience. Such consistency could provide researchers and implementers with a richer understanding of how and for whom different interventions work best and help implementers understand which behavioral interventions are more effective overall, as well as which ones might be better for a particular customer segment.

Additionally, it is important that evaluation scales include validated and standardized metrics to ensure that survey items accurately capture constructs of interest (i.e., they are measuring the actual concept of interest) and in a reliable way (i.e. that if the same person completed the measure on two different days, we would get the same answer). The process of creating and assessing the quality of variables used to measure subjective human experience is referred to as psychometrics (Kline, 2000). Psychometric testing is often comprised of the following four qualities: factor structure, reliability, validity, and sensitivity (Lewis, 1995).

This report presents psychometric testing of a set of scales that can be used to collect self-report data on of behavioral interventions as part of a comprehensive program evaluation. It builds on the behavioral intervention assessment toolkit initially developed by Karlin and Ford (2013a) and licensed under a Creative Commons Attribution 4.0 International License. This toolkit is being designed for use in assessing behavioral interventions in demand side management programs across utilities and countries.

The current project aims to refine and psychometrically validate a subset of scales from the toolkit for use in field studies within California. While the authors' eventual goal is the development of instruments that can be can be deployed more broadly across the United States and other parts of the world, we recognize that individual items in each scale may not necessarily be relevant to other regions or other cultures and some scales may need to be re-validated and re-standardized for broad implementation (see discussion of the Energy Cultures Framework in Chapter 2). Additionally, field testing is a next second and highly important step to scale validation (adding external validity to the current tests of internal validity) and the study authors are currently working to field test the scales both within and beyond California (see Chapter 4).

In Chapter 2 the development of the preliminary scale is outlined with reference to prior work and theoretical and empirical research and testing methods and results are presented. Chapter 3 presents the instrument and presents suggestions for implementation and analysis. This report concludes with a discussion addressing gaps, next steps, and suggestions for future work in Chapter 4.

2. SCALE DEVELOPMENT AND TESTING

Preliminary scales for testing were adapted from an initial "Beyond kWh" Toolkit developed by Karlin and Ford (2013a); they are based on a review of literature on knowledge, attitude, behavior, and user experience outcomes of behavioral energy interventions in demand side management programs, with an emphasis on key predictors and data collection in past research.

2.1 PREDICTORS OF ENVIRONMENTAL BEHAVIOR

A review of the literature uncovered a small number of overarching psychological theories of behavior change that have been heavily researched, and a longer list of concepts and factors (not united by any theory) that have also been empirically tested. The Theory of Planned Behavior (TPB; Ajzen, 1991) is one of the most successful models for predicting behavior. TPB classifies the beliefs guiding individuals' rational decision-making processes as: (1) behavioral beliefs (attitudes towards the behavior), (2) normative beliefs (social norms), and (3) control beliefs (perceived efficacy over the behavior). According to TPB, these three sets of beliefs influence a person's behavioral intentions, which largely determine her/his behavior. TPB has been used in hundreds of studies and has successfully predicted behaviors ranging from organ donation (Rochelau, 2013) to smoking cessation (Norman et al., 1999). It has also been used to predict environmentally responsible behavior (Karppenin, 2005; Han et al. 2010; Kim et al. 2013; Thompson & Hansen, 2013; Lam, 1999) and energy behaviors in particular (Ajzen et al., 2011; Harland et al., 1999; Lynch & Martin, 2010).

Stern's (2000) Value-Belief-Norm Theory is the second most common and researched model of environmental behavior. According to this model, values influence beliefs, which in turn motivate and guide behavior. However, these values and beliefs are only activated when something a person cares about is threatened, and that person feels that they have both the responsibility and the ability to change it. Many values can influence pro-environmental beliefs (Schultz, 2001). Biospheric values—a concern for the environment; altruistic values, such as a concern for future generations; and egoistic values of saving money or being more comfortable can all lead a person to consider engaging in pro-environmental behavior. VBN theory has successfully predicted pro-environmental behavior in a number of contexts (Chen, 2014; Guagnano, Stern, & Dietz, 1995; Sahin, 2013; Schultz & Zelezny, 1999; van Riper & Kyle, 2014; Widegren, 1998; Ziaei-Bideh & Namakshenas-Jahromi, 2014).

VBN and TPB have been directly compared and several studies (del Carmen Aguilar-Luzon et al. 2012; Kaiser et al, 2005; Lopez-Mosquera & Sanchez, 2012) have found TBP to be the better predictor of behavior (or in most cases, behavioral intention). However, it is also important to note that these approaches to predicting behavior are not mutually exclusive and that they can be integrated (e.g., Han, 2015; Turaga, Howarth, & Borsuk, 2010).

In addition to these major approaches, a number of other variables have been found to predict conservation behavior, including connection to nature (Mayer & Frantz, 2004); energy concern (Curtis, Simpson-Housley, & Drever, 1984; Verhallen & Van Raaij, 1981), price sensitivity (Long, 1993; Verhallen & Van Raaij, 1981),

environmental concern (Poortinga et al., 2003; Poortinga, 2004), and personal and social norms (Cialdini & Schultz, 2004; Goldstein, Cialdini, & Griskevicius, 2008; Nolan et al., 2008).

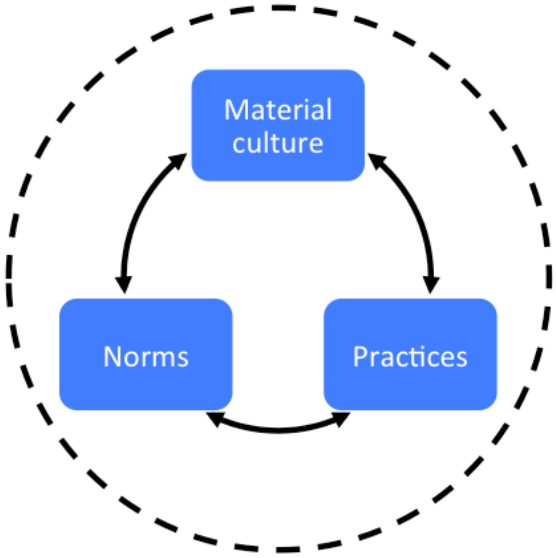
A variety of contextual variables have also been shown to predict energy conservation. Variables such as home location and size are highly related to a household's carbon footprint while variables such as income and home ownership are highly related to an individual's ability to engage in certain energy conservation behaviors (Stern, 2011). Relevant context variables identified in the literature include age (Curtis et al., 1984; Gatersleben, Steg, & Vlek, 2002; Painter, Semenik, & Belk, 1983; Sardianou, 2007), homeownership (Curtis et al., 1984; Gatersleben et al., 2002; Painter et al., 1983), income (Gatersleben et al., 2002; Ritchie, McDougall, & Claxton, 1981), education (Gatersleben et al., 2002; Painter et al., 1983), family size (Curtis et al., 1984), and home type (Sardianou, 2007). Statistical models derived from these analyses suggest that the most powerful explanation requires a combination of contextual and psychological variables.

Guagnano, Stern, and Dietz (1995) provide a useful theory that integrates psychological and contextual factors with differences in specific behaviors. Their A-B-C model posits that behavior is influenced by both attitudinal and contextual factors and that the stronger one set of factors is in predicting behavior, the less force the other exerts. If there are sufficient contextual barriers to engaging in a behavior, individuals are highly unlikely to engage in it, regardless of its alignment with self-interest or pro-social intent. For example, Black et al. (1985) found that some behaviors, such as adding home insulation, were not associated with normative beliefs when constrained by contextual factors such as household infrastructure and homeownership. On the other hand, contextual cues may trigger pro-environmental behavior, even without self-interest or pro-social motivation. Guagnano et al. (1995), for example, found that the explanatory power of personal norm beliefs decreased for recycling behavior when convenient curbside pick-up became available. Therefore, psychological variables will be most influential on pro-environmental behavior when contextual variables do not exert great influence on either promoting or restricting a behavior.

2.2 Testing Framework and Constructs

The Energy Cultures framework (Stephenson et al., 2015), depicted in Figure 1, shows how a person's material culture, practices, and norms are interlinked and embedded within a particular context that can both shape and constrain behavior. Material culture includes physical structures, building characteristics, and appliances. For example, some homes have programmable thermostats, and some do not. Practices refer to the way in which residents behave with these technologies. Even among households with programmable thermostats, some program them, and some do not. Norms refer to individuals' expectations and aspirations. Residents who are more energy conscious may be more likely to purchase and program a thermostat. Context considers how individuals are embedded in a larger social culture. For example, the technologies in the home are constrained by the technologies available in the wider culture

FIGURE 1: THE ENERGY CULTURES FRAMEWORK



Energy cultures can change over time; indeed the purpose of most interventions is to create a shift in one or more aspects of a user's energy culture. To understand this process it is important to explore all four elements of the Framework, as changes to one element (e.g. installing new technology) may result in changes in the way that householders think and act.

In addition to understanding how energy culture and energy use may change over time or as the result of participating in a program, it is important to understand how people respond to interventions (e.g. the installation of a smart thermostat), because people may not engage with information or technology that they find confusing, uninteresting, and/or unhelpful. Measuring user experience can help identify how participants respond to behavioral interventions, which may impact on the effectiveness of the program, as well as provide learning opportunities for improving program design.

Traditionally user experience is measured in laboratory settings using qualitative techniques and small sample sizes (Froehlich, 2010). While this method is useful during an iterative design phase it is less useful for evaluation of behavioral interventions, particularly when sample sizes are large. Some scales (see Karlin & Ford, 2013b, for extended discussion) have been developed for evaluating larger field studies, but these tend to be focused on software applications. The UPscale (Karlin & Ford, 2013b), builds upon decades of user experience research to create psychometrically validated scales for evaluating user perceptions of energy feedback information focusing on ease of use and engagement.

In this work the items in the original UPscale (Karlin & Ford, 2013b) have been extended to also include questions on trust, which may be an important usability construct in explaining the degree to which users believe, interact with, or act upon information provided by an intervention (Karlin, 2012; Rousseau et al., 1998). Questions related to overarching satisfaction have also been included; this will enable evaluators to explore the under-evaluated relationship between ease of use,

engagement, trust and satisfaction, and the impact that this may have on subsequent action (Flavián et al., 2006).

The tested scales combine the Energy Cultures framework (context, material culture, practices, norms) to structure the main components for use in pre- and post-intervention evaluation. In addition, the UPscale is used as the basis for evaluating user perceptions. The key constructs were developed with reference to the theoretical and empirical research, as discussed in Section 2.1, and are presented in Table 1.

TABLE 1: KEY EVALUATION CONSTRUCTS

PRE- AND POST- INTERVENTION EVALUATION ITEMS	
NORMS (PSYCHOGRAPHICS)	
Motivation to Engage	People motivated by different concerns may respond differently to the intervention
Energy Literacy	Awareness of the larger context in which behavior occurs can promote better decision making
Energy Knowledge	Knowledge specific to a particular domain is a predictor of behavior in that domain
Concern and Connection	People are more likely to take action if they feel connected to it or concerned about it
Personal Norms	Personal norms are likely to motivate behavior change when it is perceived that the environment is threatened
Social Norms	Believing that other people are engaging in/approve of a behavior is a strong predictors of that behavior
Performance Efficacy	Individuals must believe they have the ability to perform a behavior or they will not attempt to do so
Response Efficacy	Individuals must believe that the behavior will have its intended effect or they will not attempt to engage
PRACTICES (BEHAVIORS)	
Recurring Behaviors	Taking a baseline measure of behaviors that people engage in allows a precise measure of change
One-time Behaviors	
MATERIAL CULTURE	
Appliance ownership	Understanding changes in appliance ownership over time can help explain energy use changes
CONTEXT	
Physical context	Behavior is often constrained by physical and structural realities of a dwelling and its technologies
Demographics	These items can help inform who an intervention works best for, and can identify cultural constraints on change
INTERVENTION EVALUATION ITEMS	
UPscale: Ease of Use Engagement Trust	It is important to assess the subjective user experience of an energy efficiency intervention in order to understand how and why it did or did not work
Satisfaction	Understanding overarching attitudes toward the intervention is important in explaining subsequent behavior intention

The questions within each scale were developed based on a literature review conducted for the International Energy Agency Demand Side Management (IEA_DSM) Programme Task 24 on Behavior Change (Karlin & Ford, 2013a) (Table 2). From this literature, sets of similar or related questions were identified under each of the key constructs of norms, practices, material culture, context, and user experience, and used to develop the set of potential survey questions validated in this study[1].

TABLE 2: SURVEY LITERATURE

REVISED INSTRUMENT	CITATIONS
MATERIAL CULTURE	
Appliance ownership	Wooliscroft, B. (2015). National Household Survey of Energy and Transportation: Energy Cultures Two. Centre for Sustainability, University of Otago. ISBN: 978-0-9941219-0-5
Programmable/ smart thermostat ownership	Allcott, H., & Rogers, T. (2012). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. NBER Working Paper Series.
PRACTICES	
One time behaviors (and intentions)	Rambo, E., & Feldman, S. (2003). What is it I need to know? The Relationship Between Information Seeking and Intended Action Relating to Energy Efficiency. In energy program evaluation conference (pp. 469–480). Seattle.
	Staats, H., Harland, P., & Wilke, H. a. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. Environment & Behavior, 36(3), 341–367. doi:10.1177/0013916503260163
	Allcott, H., & Rogers, T. (2012). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. NBER Working Paper Series.
	Rambo, E., & Feldman, S. (2003). What is it I need to know? The Relationship Between Information Seeking and Intended Action Relating to Energy Efficiency. In energy program evaluation conference (pp. 469–480). Seattle.
	Feldman, S., & Rambo, E. (2003). How am I Doing? Tracking the Effectiveness of Advertising an Energy-Efficiency Program. In Energy Program Evaluation Conference (pp. 403–416).
	Jackson, C., Peters, J., Spahic, M., & Lutzenhiser, S. (2009). Trends in ENERGY STAR ® Awareness: Results from Four National Surveys, 2002 - 2008. In energy program evaluation conference (pp. 382–393).
	Lynch, D., & Martin, P. (2010). How energy efficiency programs influence energy use: an application of the theory of planned behaviour. In ECEEE Summer Study Proceedings (pp. 2037–2048).
	Nishio, K., & Ofuji, K. (2012). Behavior Change and Driving Forces to Save Electricity in the Electricity Crisis in Japan. In International Energy Program Evaluation (pp. 1–12).
	Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect

[1] Further information related to prior work and previously used instruments can be found in Karlin & Ford, 2013a and Karlin et al., 2015. The full question set with sources is available from the authors upon request - contact bkarlin@seechangeinstitute.com or rford@seechangeinstitute.com.

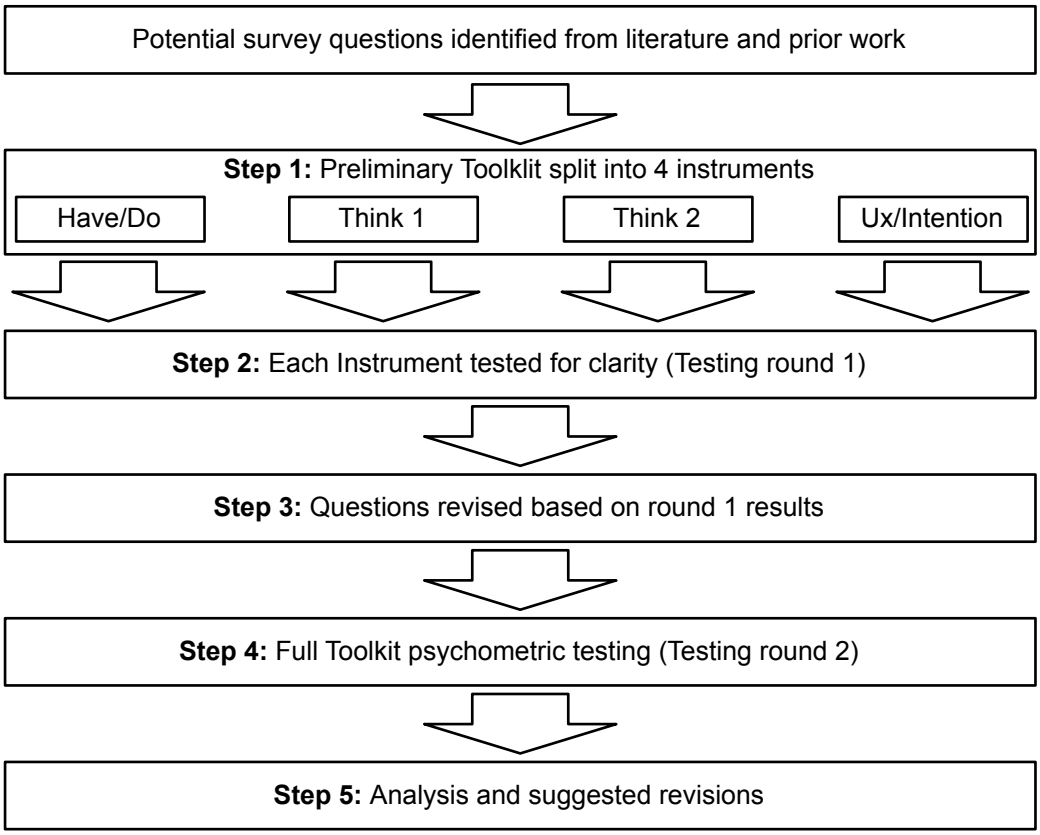
	of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. Journal of Environmental Psychology, 27(4), 265-276.
	Benders, R. M., Kok, R., Moll, H. C., Wiersma, G., & Noorman, K. J. (2006). New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options. Energy policy, 34(18), 3612-3622.
	Tiedemann, K., Sulyma, I., & Rebman, M. (2007). Measuring the Impact of Time of Use Rates on Peak and Off-peak Energy Consumption: Some Results from a Randomized Controlled Experiment. In 2007 Energy Program Evaluation Conference (pp. 77–87). Chicago.
	Wooliscroft, B. (2015). National Household Survey of Energy and Transportation: Energy Cultures Two. Centre for Sustainability, University of Otago. ISBN: 978-0-9941219-0-5
Recurring behaviors (and intentions)	Rambo, E., & Feldman, S. (2003). What is it I need to know? The Relationship Between Information Seeking and Intended Action Relating to Energy Efficiency. In energy program evaluation conference (pp. 469–480). Seattle.
	Staats, H., Harland, P., & Wilke, H. a. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. Environment & Behavior, 36(3), 341–367. doi:10.1177/0013916503260163
	Benders, R. M., Kok, R., Moll, H. C., Wiersma, G., & Noorman, K. J. (2006). New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options. Energy policy, 34(18), 3612-3622.
	Tiedemann, K., Sulyma, I., & Rebman, M. (2007). Measuring the Impact of Time of Use Rates on Peak and Off-peak Energy Consumption: Some Results from a Randomized Controlled Experiment. In 2007 Energy Progran Evaluation Conference (pp. 77–87). Chicago.
	Allcott, H., & Rogers, T. (2012). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. NBER Working Paper Series.
	Nishio, K., & Ofuji, K. (2012). Behavior Change and Driving Forces to Save Electricity in the Electricity Crisis in Japan. In International Energy Program Evaluation (pp. 1–12).
	Kurz, T., Donaghue, N., & Walker, I. (2005). Utilizing a social-ecological framework to promote water and energy conservation: A field experiment. Journal of Applied Social Psychology, 35(6), 1281-1300. doi:http://dx.doi.org/10.1111/j.1559-1816.2005.tb02171.x
	Tiedemann, K., Sulyma, I., & Rebman, M. (2007). Measuring the Impact of Time of Use Rates on Peak and Off-peak Energy Consumption: Some Results from a Randomized Controlled Experiment. In 2007 Energy Program Evaluation Conference (pp. 77–87). Chicago.
	Fonseca, S., & Nave, J. G. (2009). From structural factors to individual practices: reasoning on the main paths for action on energy efficiency. In ECEEE 2009 summer study (pp. 1865–1873).
	Freeman, D. J., & Skumatz, L. A. (2012). Widgets versus Actions: Measuring the Role of Behavior Change in DSM Programs. In international energy program evaluation conference (pp. 1–12).
NORMS	
Connection and concern	Mayer, F. S., & Frantz, C. M. (2004). The Connectedness to Nature Scale: A Measure of Individuals' Feeling in Community with Nature. *Journal of Environmental Psychology, 24,* 504-515.
	Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use

	and savings? Journal of Economic Psychology, 30(5), 711-720. doi:http://dx.doi.org/10.1016/j.joep.2009.05.006
	Petersen, J. E., Frantz, C. M., Shammin, Md. R., Yanisch, T., Tincknell, E., & Myers, N. (2015). Electricity and water use savings on college and university campuses in response to national competitions among dormitories: quantifying relationships between behavior, strategies and psychological metrics. PLOS ONE.
	Davis, A. & Stroink, M. (2015). The Relationship between Systems Thinking and the New Ecological Paradigm. Systems Research and Behavioral Science.
Energy Literacy	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. Buildings, 1–12.
	Nishio, K., & Ofuji, K. (2012). Behavior Change and Driving Forces to Save Electricity in the Electricity Crisis in Japan. In International Energy Program Evaluation (pp. 1–12).
Energy Knowledge	Nishio, K., & Ofuji, K. (2012). Behavior Change and Driving Forces to Save Electricity in the Electricity Crisis in Japan. In International Energy Program Evaluation (pp. 1–12).
	Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. Journal of Environmental Psychology, 27(4), 265-276.
	Fonseca, S., & Nave, J. G. (2009). From structural factors to individual practices: reasoning on the main paths for action on energy efficiency. In ECEEE 2009 summer study (pp. 1865–1873).
Personal Norms	Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? Journal of Economic Psychology, 30(5), 711-720. doi:http://dx.doi.org/10.1016/j.joep.2009.05.015
	Lynch, D., & Martin, P. (2010). How energy efficiency programs influence energy use: an application of the theory of planned behaviour. In ECEEE Summer Study Proceedings (pp. 2037–2048).
Social Norms	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), 913–23. doi:10.1177/0146167208316691
	Lynch, D., & Martin, P. (2010). How energy efficiency programs influence energy use: an application of the theory of planned behaviour. In ECEEE Summer Study Proceedings (pp. 2037–2048).
Performance Efficacy	Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? Journal of Economic Psychology, 30(5), 711-720. doi:http://dx.doi.org/10.1016/j.joep.2009.05.011
	Rambo, E., & Feldman, S. (2003). What is it I need to know? The Relationship Between Information Seeking and Intended Action Relating to Energy Efficiency. In energy program evaluation conference (pp. 469–480). Seattle.
	Lynch, D., & Martin, P. (2010). How energy efficiency programs influence energy us: an application of the theory of planned behaviour. In ECEEE Summer Study Proceedings (pp. 2037–2048).
	Mizobuchi, K., & Takeuchi, K. (2013). The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in Japan, 63, 775–787. doi:10.1016/j.enpol.2013.08.064
Response Efficacy	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), 913–23. doi:10.1177/0146167208316691
	Abrahamse, W., & Steg, L. (2009). How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? Journal of Economic Psychology, 30(5), 711-720. doi:http://dx.doi.org/10.1016/j.joep.2009.05.007
	Fonseca, S., & Nave, J. G. (2009). From structural factors to individual practices: reasoning on the main paths for action on energy efficiency. In ECEEE 2009 summer study (pp. 1865–1873).
Motivation to Engage	Erickson, T., Li, M., Kim, Y., Deshpande, A., Sahu, S., Chao, T., … Naphade, M. (2013). The Dubuque Electricity Portal: Evaluation of a City-Scale Residential Electricity Consumption Feedback System. In CHI 2013 (pp. 1203–1212). Paris, France.
	Moran, D., & Forster, H. (2012). You Want Me to Do What? Smart Grid and Demand Response Pilots Test the Waters with Residential Customers The Pilots: Similarities and Differences, 186–197.
	Nishio, K., & Ofuji, K. (2012). Behavior Change and Driving Forces to Save Electricity in the Electricity Crisis in Japan. In International Energy Program Evaluation (pp. 1–12).
	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. Buildings, 1–12.
	Stamminger, P. R., & Anstett, V. (2013). Effectiveness of demand side management by variable energy tariffs in the households – results of an experimental design with a fictive tariff model (pp. 2159–2166). Retrieved from http://proceedings.eceee.org/papers/proceedings2013/8-020-13_Stamminger.pdf?returnurl=http://proceedings.eceee.org/visabstrakt.php?event=3&doc=8-020-13
	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), 913–23. doi:10.1177/0146167208316691
	Staats, H., Harland, P., & Wilke, H. a. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. Environment & Behavior, 36(3), 341–367. doi:10.1177/0013916503260163
INTERVENTION EVALUATION ITEMS (UPSCALE)	Karlin, B., & Ford, R. (2013). The Usability Perception Scale (UPscale): A measure for evaluating feedback displays. Proceedings of the 2013 Human Computer Interaction Conference, (2013).
	Moere, A. Vande, Tomitsch, M., Hoinkis, M., Trefz, E., Johansen, S., & Jones, A. (2011). Comparative Feedback in the Street: Exposing Residential Energy Consumption on House Façades, 470–488.
	Moran, D., & Forster, H. (2012). You Want Me to Do What? Smart Grid and Demand Response Pilots Test the Waters with Residential Customers The Pilots: Similarities and Differences, 186–197.
	Benders, R. M., Kok, R., Moll, H. C., Wiersma, G., & Noorman, K. J. (2006). New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options. Energy policy, 34(18), 3612-3622.
	Staats, H., Harland, P., & Wilke, H. a. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. Environment & Behavior, 36(3), 341–367. doi:10.1177/0013916503260163

The preliminary survey instrument was evaluated in two rounds of testing (see Figure 2). First, feedback was solicited on the entire set of items to ensure that each question was worded clearly and had appropriate response options. Based on this feedback from participants minor revisions were made to many of the questions. Next, the entire set of questions was administered to test for psychometric properties (factor structure, reliability, convergent and divergent validity, and predictive validity) with a goal of reducing the number of items and procuring psychometrically validated scales.

FIGURE 2: TESTING PROCESS



2.3 TESTING ROUND 1

PARTICIPANTS

Eighty participants were recruited from Mechanical Turk (MTurk) to complete a survey about energy use and were paid $.35 for their time; 20 participants each completed a subset of the items considered for inclusion in the Toolkit. Fifty-seven (71.25%) lived in houses, 21 (26.25%) in an apartment, 1(1.25%) stated "Other", and 1 (1.25%) did not answer the question. A majority of the participants (59%) owned their home. The mean number of people living in the home (only those that spend more than 50% of their time at the home) was 2.7, with a standard deviation of 1.35. Most of the participants (54%) ranged in ages 25 to 45. Out of the 80

participants, 34 (42%) were female and 46 (58%) were male. Sixty-nine (86%) of the participants identified as White, 5(6%) identified as Hispanic, 1(1%) identified as Black, and 3 (4%) identified themselves as more than one category.

METHODS

For initial testing of the Toolkit, questions were split into four smaller surveys (see Appendices A – D). Each sub-survey was developed to have similar completion times. The first asked questions about the respondent's material culture and energy practices (i.e. what they *have* and the things they *do*). The second and third sub-surveys asked questions about norms, i.e. what respondent's *think*. The final sub-survey included questions about usability and behavioral intention. All four included the same context questions.

The surveys were constructed in SurveyMonkey and participants were recruited via Amazon Mechanical Turk (see Appendix E for posting). To test the question clarity, the following mandatory open-ended prompt was added after each question:

Before going on to the next page, please take a moment now to review the questions above. Was the wording clear or confusing? Did you have trouble answering any question? Were you provided with an answer choice that made sense? Please make any notes or observations in the text box below.

In addition, the following "trick" question was added into each of the surveys to ensure that participants were reading survey questions thoroughly;

In order to facilitate our research, we are interested in knowing certain factors about you. Specifically, we are interested in whether you actually take the time to read the directions; if not, then the data we collect based on your responses will be invalid. So, in order to demonstrate that you have read the instructions, please ignore the next question (leave all of the answer options unchecked) and write the sentence "I read the instructions" in the textbox below the prompt for additional comments.

How easy do you think it is for the average American to save energy?

- *Very easy*
- *Somewhat easy*
- *Neutral*
- *Somewhat difficult*
- *Very difficult*

Please use the space below if you have any additional comments.

Participants who failed to correctly answer this question (N = 10) were removed from the data set.

RESULTS

Reviewing feedback on questions, it appeared that most were clear and understandable to participants. Comments that reflected on the clarity of the questions asked and indicated that an amendment was required were reviewed by the research team and used to revise the instruments. Following this revision, the scales were collapsed back into a single survey for a second round of testing.

2.4 TESTING ROUND 2

PARTICIPANTS

Participants were recruited from MTurk to participate in a survey about energy use and were paid $2 for their time. As in Round 1, a trick question was included to determine whether participants read the directions; those who failed to correctly answer this question (N = 47) were removed.

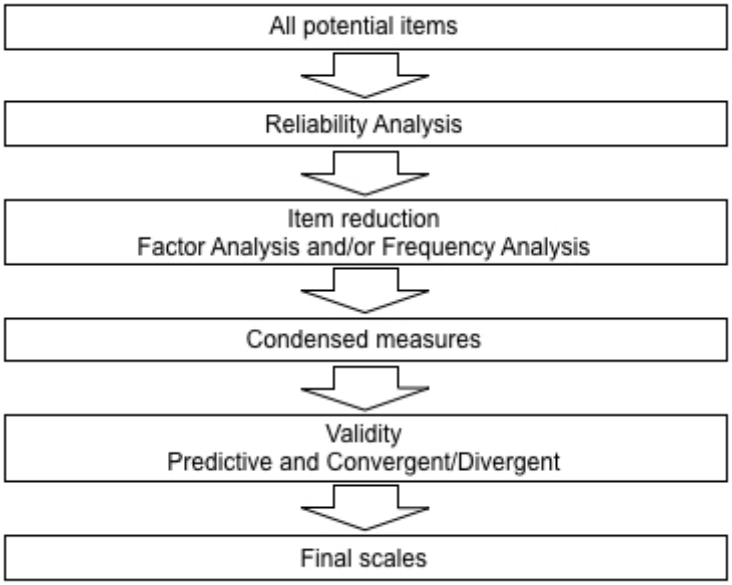
Of the 520 final participants, most participants 349 (67.1%) lived in houses; 138 (26.5%) lived in an apartment, 22 (4.2%) in a manufactured or mobile home, 7(1.3%) stated "Other", and 3 (.6%) did not answer the question. Results also showed that 249 (47.9%) participants owned their home, while 248 (47.7%) rented their home. The mean number of people living in the home (only those that spend more than 50% of their time at the home) was 2.8, with a standard deviation of 1.35.

Most of the participants 308 (59.2%) ranged from 25 to 45 years. Out of the 520, 261 (50.2%) were female, 253 (48.7%) were male, and 3 (.6%) did not respond. Four hundred thirteen (79.4%) of the participants identified themselves as White, 37 (7.1%) identified themselves as Hispanic, 43 (8.3%) identified as Black, 8 (1.5%) identified as American Indian or Alaskan Native, 43 (8.3%) identified as Asian, 1(.2%) participant identified as Middle Eastern or North African, 1(.2%) participant identified as Native Hawaiian or other Pacific Islander, and 4 (.8%) identified themselves as "Other".

METHODS

Participants completed the entire set of potential Toolkit items (including the modifications based on Round 1 results) as well as additional scales included to establish convergent and divergent validity (see Appendix F). The survey was constructed in SurveyMonkey then put on MTurk for testing (see Appendix G for posting). Analysis of survey results was undertaken to identify a condensed set of items that reliably measure each construct, as illustrated in Figure 3.

FIGURE 3: ANALYSIS PROCESS



Final scales were developed through: (1) ensuring items designed to measure a single construct have high reliability; (2) reducing items to create shorter scales which may be more useful for evaluators; (3) establishing convergent and divergent validity; and (4) testing each construct's predictive validity (using self reported behavior and behavioral intentions).

For scales measuring latent constructs, factor analysis was used to reduce the number of items such that repetitive questions are avoided in the final scale. For those items not designed to measure latent constructs (i.e. knowledge, behavior, and behavioral intention), inspection of frequencies was used to identify items with extremely low variability to create scales. The scales resulting from item reduction (either by factor analysis or frequency analysis) were subjected to convergent and divergent validity analyses. Finally, the ability of the scales to predict behavior and behavioral intention was tested using a series of regression equations.

RESULTS

After rescoring reverse-worded items, reliability was computed for each set of questions designed to measure a single construct using Cronbach's alpha, a summary statistic of how well individual items correlate with each other. Cronbach's alphas were also calculated on the scales that were used to establish convergent and divergent validity. All scales had alphas ranging from acceptable (.60) to excellent (.90) (see Table 2). There were no items that substantially decreased the reliability of any scale, so no items were identified as problematic based on this criterion. Initial scales were computed by averaging all items together. The means and standard deviations for all scales appear in Table 3.

TABLE 3: RELIABILITIES OF TOOLKIT SCALES AND CONVERGENT/DIVERGENT VALIDITY SCALES

SCALE NAME	# ITEMS	N	ALPHA
PRE- AND POST- INTERVENTION EVALUATION ITEMS			
NORMS (PSYCHOGRAPHICS)			
Motivation to Engage[1]			
Energy Literacy -- Awareness	4	504	0.865
Energy Knowledge[1]			
Connection and Concern	10	502	0.892
Personal Norms	8	500	0.897
Social Norms	6	507	0.616
Performance Efficacy	9	501	0.852
Response Efficacy	8	504	0.874
PRACTICES (BEHAVIORS)			
Recurring Behaviors[1]			
One-time Behaviors[1]			
MATERIAL CULTURE			
Appliance Ownership[1]			
CONTEXT			
Physical Context[1]			
Demographics[1]			
Utility Questions[2]	9	385	0.425
INTERVENTION EVALUATION ITEMS			
UPscale - Ease of Use	7	499	0.870
UPscale - Engagement	9	497	0.90
UPscale - Trust	2	509	0.924
Satisfaction	2	513	0.87
SCALES FOR CONVERGENT/DIVERGENT VALIDITY			
New Environmental Paradigm	15	485	0.907
Connectedness to Nature	14	486	0.940
Environmental Attitudes -- Internal	7	520	0.874
Environmental Attitudes -- External	7	520	0.634
NEO Openness	8	520	0.738
NEO Conscientiousness	8	520	0.738
NEO Extraversion	8	520	0.748
NEO Agreeableness	8	520	0.696
NEO Neuroticism	8	520	0.705
Frugality	8	502	0.881

[1] Chronbach alphas are not reported on Motivation to Engage, Energy Knowledge, or any Practices, Material Culture, or Context questions because they are not sets of questions designed to assess a single underlying construct.

[2] Utility questions were not uncovered during the literature review and thus have not been discussed in earlier sections of this report; they were provided by SCE to be included in testing.

TABLE 4: DESCRIPTIVE STATISTICS OF RESPONSES TO EACH MAIN CONSTRUCT

SCALE	N	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
PRE- AND POST- INTERVENTION EVALUATION ITEMS					
NORMS (PSYCHOGRAPHICS)					
Motivation to Engage[1]					
Energy Literacy -- Awareness	516	1.00	5.00	3.4724	0.85166
Energy Knowledge	516	0.00	1.00	0.6256	0.19769
Connection and Concern	515	1.11	5.00	3.7250	0.72524
Personal Norms	511	1.50	5.00	3.8430	0.72241
Social Norms	513	1.50	4.50	3.0705	0.58118
Performance Efficacy	515	1.44	5.00	3.8190	0.62014
Response Efficacy	514	1.75	5.00	3.8460	0.67604
PRACTICES (BEHAVIORS)					
Recurring behaviors	514	1.33	5.00	3.4427	0.61613
Behavioral Intention – Recurring behaviors	513	1.29	4.93	3.7020	0.56933
One-time behaviors	517	0.00	29.00	5.7950	4.96934
Behavioral Intention – One-time behaviors	517	0.00	20.00	1.3095	3.14522
MATERIAL CULTURE					
Appliance ownership[1]					
CONTEXT					
Physical context[1]					
Demographics[1]					
Utility questions	515	1.00	4.63	3.1944	0.51136
INTERVENTION EVALUATION ITEMS					
UPscale - Ease of Use	514	2.14	5.00	4.1990	0.65349
UPscale - Engagement	514	1.00	5.00	3.9741	0.67206
UPscale - Trust	514	1.00	5.00	3.9377	0.76537
Satisfaction	514	1.00	5.00	4.0914	0.70324
SCALES FOR CONVERGENT/DIVERGENT VALIDITY					
New Environmental Paradigm	515	1.60	5.00	3.5689	0.71210
Connectedness to Nature Scale	514	1.00	5.00	3.4754	0.73027
Environmental Attitudes -- Internal	517	1.57	5.00	3.8663	0.75908
Environmental Attitudes -- External	517	1.00	4.43	2.6259	0.56698
NEO Openness	517	0.00	1.00	0.6310	0.26020
NEO Conscientiousness	517	0.00	1.00	0.7004	0.24223
NEO Extraversion	517	0.00	0.88	0.3426	0.27827
NEO Agreeableness	517	0.00	1.00	0.8027	0.20354
NEO Neuroticism	517	0.00	1.00	0.4717	0.22370
Frugality	513	2.38	5.00	4.1909	0.53338

[1] Means and SDs are not included for these items as they are categorical variables.

Item Reduction: Factor Analysis

Factor analysis (using a non-orthogonal rotation and factor extraction based on eigenvalues over 1) was conducted on items designed to measure motivation to engage, energy literacy, connection and concern, personal norms, social norms, performance efficacy, response efficacy, and user experience to test whether the items designed to assess these constructs cohered along a single dimension. Where appropriate, factor analysis was run combining all related constructs (e.g. performance and response efficacy, personal norms and social norms, subscales of the UPscale) to demonstrate that these concepts, though related, are in fact distinct.

Factor analysis was not run on the established scales included for convergent and divergent validity, as these scales have already been fully validated. Factor analysis was also not run on the energy knowledge questions or behavior questions, as these are not items designed to measure the same latent construct. Although also not designed to measure a latent construct, we did factor analyze the utility questions and the motivation to engage questions to determine whether common themes emerged. A summary of the findings is shown in Table 4.

Table 4: Factor analysis findings

Construct	# factors	% variance explained
Norms (Psychographics)		
Motivation to engage	3	0.64
Energy Literacy -- Awareness	1	0.72
Concern and Connection	2	0.64
Norms (personal, social)	2	0.54
Efficacy (performance, response)	2	0.52
Utility Questions	3	0.55
Upscale (ease of use, engagement, trust)	3	0.65

Factor loadings were used to identify and validate subscales within each construct; they were also used to identify the items that best represented (or "loaded" on) each factor, and to eliminate items that did not load well on to any factor, enabling condensed scales to be created. When items loaded equally well, an effort to include reverse-scored items was made. Factor loadings for each construct are presented in Appendix I.

While factor analysis enabled most of the constructs to be evaluated using condensed scales with just 2-3 items, item reduction was not performed on the UPscale subscales of ease of use or engagement (the trust subscale only contained 2 items to start with so both of these were retained). The UPscale questions are designed to measure a participant's subjective response to an intervention, unlike all the other items, which are designed to measure various constructs at any point in time. Because testing to date has been based on hypothetical scenarios rather than under field study settings, we decided that it would be irresponsible to remove items and condense the UPscale subscales scales until full field testing of this construct has been undertaken.

Reliabilities for each scale were calculated. All scales had good to excellent reliability with the exception of the utility subscales (which again were not designed to measure a latent construct). (Cronbach's alpha ranged from .738-.924), see Table 5.

TABLE 5: RELIABILITIES OF CONDENSED SCALES

CONSTRUCT	# ITEMS	N	ALPHA
PRE- AND POST- INTERVENTION EVALUATION ITEMS			
NORMS (PSYCHOGRAPHICS)			
Motivation – cost[2]	1		
Motivation - prosocial	3	514	0.801
Motivation - self comfort	3	506	0.796
Energy Literacy – Awareness	2	505	0.861
Connection	2	514	0.769
Concern	2	510	0.843
Personal Norms	3	504	0.808
Social Norms	2	510	0.738
Performance Efficacy	2	510	0.820
Response Efficacy	3	509	0.828
PRACTICES (BEHAVIORS)			
Recurring Behaviors[1]			
One-time Behaviors[1]			
MATERIAL CULTURE			
Appliance ownership[1]			
CONTEXT			
Physical context[1]			
Demographics[1]			
Utility Questions: Response to peak demand	4	455	0.655
Utility Questions: Anti-programming	3	468	0.443
Utility Questions: Utility perceptions	2	440	0.505
INTERVENTION EVALUATION ITEMS			
UPscale - Ease of Use	7	499	0.870
UPscale - Engagement	9	497	0.900
UPscale - Trust	2	509	0.924
Satisfaction	2	513	0.870

[1] Alpha is not reported for these questions they are not designed to assess a single underlying construct

[2] Motivation -- cost was a single item, thus reliabilities could not be computed

In summary, factor analyses revealed that the vast majority of the original individual items loaded in ways that were consistent with the conceptual variables they were intended to measure. For each construct, two to four items were identified that loaded strongly on the main factor and did not load on other factors (Appendix I presents the factor loadings; Appendix H presents all original items and indicates which were retained). 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.

Item Reduction: Frequency Analysis

Factor analysis is not appropriate for those items not designed to measure latent constructs, and instead, an inspection of frequencies can be used for item reduction and the creation of condensed scales. Questions to which nearly all participants give the same answer do not provide much useful information, and when the goal is to keep the number of questions to a minimum, these items are candidates for elimination. The material culture, energy literacy -- knowledge, behavior (both one-time and recurring), and behavioral intention questions were examined to determine whether any items showed extreme lack of variability (see Table 6).

All material culture, behavior, and behavioral intention items showed substantial variability (no more than 80% of participants selected a single answer; 80% indicated they have and use a washing machine, and 77% indicated they have and use a dryer). This suggests that each of these questions provides useful information: one cannot assume the answer to the question a priori based on the ubiquitous presence of the behavior or appliance. There may, of course, be other reasons to eliminate items (some behaviors or appliances may be irrelevant to the goals of a particular study).

Several of the knowledge items showed restricted variability, however, with over 80% of participants choosing the correct answer. In particular, nearly all participants answered Question 2 (Ceiling insulation can help keep your home warm in winter and cool in summer) and Question 5 (Washing clothes at lower temperatures uses less energy than higher temperatures) correctly. These items are candidates for being dropped in future iterations.

Table 6: Frequencies for energy knowledge questions

Questions	N	Agree	Disagree	I don't know
Turning the thermostat up higher will make the room get warmer more quickly	517	48.4%	42.0%	9.5%
Ceiling insulation can help keep your home warm in winter and cool in summer	517	89.7%	1.4%	8.5%
Reducing my water usage does not save energy	517	7.5%	82.6%	9.1%
Energy efficient light bulbs use less than half of the electricity of incandescent bulbs	517	81.8%	4.4%	13.0%
Washing clothes at lower temperatures uses less energy than higher temperatures	517	86.5%	1.7%	11.6%
Leaving a window open on an upper floor lets heat escape on a hot day	517	60.0%	15.1%	24.8%
The refrigerator uses more electricity than the air conditioner	517	25.9%	38.3%	35.0%

Convergent and Divergent Validity

To establish both convergent and divergent validity of the items, all scales were compared to a set of previously validated scales: (1) Connectedness to Nature (CNS), (2) New Environmental Paradigm (NEP), (3) Environmental Attitudes Scale (EAS) including both Internal and External motivation subscales (EAS-I and EAS-E respectively), (4) Frugality Scale, and the (5) Big Five Personality Inventory (extraversion, neuroticism, openness, agreeableness, conscientiousness). The EAS-E served as a measure of domain-specific socially desirable responding. These scales can be found in Appendix J.

Convergent validity is demonstrated when items correlate with other constructs that should theoretically be related. Divergent validity is established when items do not correlate with constructs that should be theoretically distinct. The following a priori predictions were made:

- Energy awareness would correlate strongly with CNS, NEP, EAS-I, frugality, and conscientiousness. It would not correlate strongly with other personality measures, particularly extraversion and neuroticism.
- Energy knowledge was expected to correlate with conscientiousness and perhaps frugality, but with very little else.
- Personal norms would correlate with CNS, NEP, EAS-I, conscientiousness, and agreeableness. It would not correlate strongly with other personality measures, particularly extraversion and neuroticism.
- Social norms would correlate most strongly with EAS-E, with agreeableness, and perhaps with extraversion. It would not correlate strongly with other personality measures.
- Performance efficacy would correlate with CNS, NEP, EAS-I, and conscientiousness. It would not correlate strongly with other personality measures, particularly extraversion and neuroticism.
- Response efficacy would correlate with CNS, NEP, EAS-I, agreeableness and conscientiousness. It would not correlate strongly with other personality measures, particularly neuroticism.
- Engagement should correlate with openness and conscientiousness; ease of use should correlate with conscientiousness; trust should correlate with agreeableness and openness.
- Prosocial motivation should correlate with NEP, CNS, EAS-I, openness, and agreeableness, but not frugality, extraversion and neuroticism.
- Self-comfort motivation should correlate negatively with NEP, CNS, and EAS-I, and positively with neuroticism.
- Cost motivation should correlate with frugality and conscientiousness, but little else.
- EAS-External serves as a measure of socially desirability; to the extent that our measures are not primarily measuring social desirability, correlations with EAS-E should be weaker than correlations with the CNS, NEP and EAS-I.

Correlations indicated that vast majority of these predictions were supported (see Appendix K). Exceptions include: knowledge did not correlate with conscientiousness; prosocial motivation did correlate with frugality and extraversion; self-comfort motivation did not correlate at all with NEP, CNS, and EAS-I; cost motivation surprisingly correlated with NEP, CNS, and EAS-I; engagement also correlated with agreeableness. Engagement, ease of use, and trust all correlated with NEP, CNS, and EAS-I, perhaps because people high on these traits found the information presented interesting and familiar.

Particular attention was paid to the relationship between the scales developed here and EAS-E, a measure of the tendency to engage in socially desirable responding about environmental concern (see Table 7). The EAS-E did not correlate with self-comfort motivation, cost-motivation, engagement, or trust. The EAS-E was correlated significantly with connection, concern, energy awareness, personal norms, social norms, response efficacy, prosocial motivation, and ease of use (r's ranging from .15 to .39). However in all cases except social norms (as predicted), the correlation between EAS-I and the variable was much stronger (usually 2 to 3 times stronger) than the correlation between EAS-E and the variable. This suggests that internal motivations for pro-environmental behavior are more predictive of responses to survey questions than motivation to appear environmental.

TABLE 7: CORRELATIONS WITH INTERNAL AND EXTERNAL PRO-ENVIRONMENTAL MOTIVATION

SCALE	EAS-I	EAS-E
PRE- AND POST- INTERVENTION EVALUATION ITEMS		
NORMS (PSYCHOGRAPHICS)		
Motivation - prosocial	.451**	.388**
Motivation - self comfort	-0.033	-0.06
Motivation - cost	.258**	0.023
Energy Literacy -- Awareness	.421**	.273**
Energy Knowledge[1]		
Connection	.540**	.356**
Concern	.570**	.280**
Personal norms	.701**	.297**
Social norms	0.026	.294**
Performance efficacy	.390**	.150**
Response efficacy	.560**	.237**
PRACTICES (BEHAVIORS)		
Recurring Behaviors[1]		
One-time Behaviors[1]		
MATERIAL CULTURE		
Appliance ownership[1]		
CONTEXT		
Physical context[1]		
Demographics[1]		
INTERVENTION EVALUATION ITEMS (UPSCALE)		
UPscale - Ease of use	-.430**	-.234**
UPscale - Engagement	.428**	0.053
UPscale - Trust	.254**	0.053
Satisfaction	.255	.010

** Correlation is significant at the 0.01 level (2-tailed).

[1] Correlations are not reported because these variables are not continuous

Finally, each energy knowledge question, as well as the average of all energy knowledge questions, were correlated with self-reported behavior and behavioral intentions (both one-time and recurring). These correlations are presented in Table 8. While the literature suggests that general knowledge does not predict behavior, knowledge specific to the domain in question should be related to behavior in that domain. Knowledge was a comparatively good predictor of recurring behaviors and behavioral intentions (though none of the correlations were particularly high). These questions also significantly predicted the number of one-time behaviors respondents had already completed, suggesting that the knowledge questions included in the toolkit are in fact relevant to behaviors of interest. The correlations between knowledge and the intention to perform one-time behaviors were not positive. Those who did not plan to do these behaviors were both people who had already completed them and people who had no intention of completing them (perhaps because of constraints in their current living situation, such as renting), making an interpretation of the correlation difficult.

TABLE 8: CORRELATIONS BETWEEN ENERGY KNOWLEDGE ITEMS AND BEHAVIORS/BEHAVIORAL INTENTIONS

	ONE TIME BEHAVIORS	RECURRING BEHAVIORS	ONE TIME BEHAVIORAL INTENTIONS	RECURRING BEHAVIORAL INTENTIONS
Energy knowledge average	.165**	.201**	-.105*	.228**
Turning the thermostat up higher will make the room get warmer more quickly	.123**	.127**	-0.071	.088*
Ceiling insulation can help keep your home warm in winter and cool in summer	.141**	.102*	-0.083	.149**
Reducing my water usage does not save energy	.131**	.243**	-.144**	.283**
Energy efficient light bulbs use less than half of the electricity of incandescent bulbs	.098*	.145**	0.012	.177**
Washing clothes at lower temperatures uses less energy than higher temperatures	.099*	.173**	-0.053	.158**
Leaving a window open on an upper floor lets heat escape on a hot day	-0.015	-.106*	-0.011	-0.05
The refrigerator uses more electricity than the air conditioner	0.022	0.048	-0.028	0.05

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

In summary, the pattern of correlations suggests that both the original items and the condensed scales relate to theoretically relevant variables in predicted ways. By and large, they correlate with variables they should be related to, and are unrelated to variables that should be distinct. Social desirability is always a concern, but results suggest that social desirability is not the main driver behind respondents' answers, as internal environmental attitudes (as measured by the EAS-I) are a stronger predictor.

Predictive Validity

The final test of the scale is whether the items predict behavioral intentions and behavior. Testing in a field setting is essential, but the behavior and behavioral intentions data collected in this study allow for a preliminary test of predictive validity. Four dependent variables were used to conduct these analyses:

1. The total number of one-time energy saving behaviors participants reported having already taken;
2. The average frequency with which participants reported engaging in recurring energy saving behavior;
3. The total number of one-time behaviors participants said they plan to do; and
4. The average rated likelihood of engaging in the recurring behaviors in the next six months.

The Theory of Planned Behavior (Ajzen, 1991) and the Value Belief Norm Theory (Stern, 2000) were used to guide predictive validity testing. VBN posits that values lead one to become aware of threats to 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 behavior, in this case energy reducing behavior.

The TPB predicts that social norms, attitudes, and efficacy all contribute to forming behavioral intentions, which in turn predict behavior. Because of the cross-sectional nature of the data set the relationship between intention and behavior could not be tested. However, the first stage of the model could be. The condensed scales measuring social norms, personal norms (attitudes), and both performance and response efficacy were used to predict intention to engage in recurring conservation behaviors and intention to engage in one-time conservation behaviors. As predicted, all four factors predicted intention to engage in recurring conservation behaviors. All but performance efficacy significantly predicted intention to engage in one-time behaviors; the effect was in the right direction but not significant.

Table 9: Regression equations testing the Theory of Planned Behavior

DV	IV	B	P
Predictors of intention to perform recurring behaviors	Personal norms	0.149	<.001
	Social norms	0.094	0.012
	Performance efficacy	0.180	<.001
	Response efficacy	0.221	<.001
Predictors of intention to perform one-time behaviors	Personal norms	0.461	0.089
	Social norms	0.867	0.001
	Performance efficacy	-0.417	0.133
	Response efficacy	-0.761	0.034

To test VBN theory, four sets of path analyses were conducted with the condensed scales, using both behavior variables and both behavioral intention variables as the final predictors in the model. As predicted by the model, pro-social values and cost concerns both predicted concern, which in turn predicted personal norms. Personal

norms did not significantly predict intention to engage in one-time conservation behaviors, but it significantly predicted one-time behaviors already completed, frequency of engaging in recurring conservation behaviors, and intention to engage in recurring conservation behaviors.

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

TABLE 10: REGRESSION EQUATIONS TESTING THE VALUE BELIEF NORM THEORY

DV	IV	B	P
Concern	Prosocial motivation	0.563	<.001
	Comfort motivation	-0.026	0.657
	Cost motivation	0.141	0.009
Personal norms	Concern	0.564	<.001
One-time behavioral intentions	Personal norms	0.036	0.834
Recurring behavioral intentions	Personal norms	0.972	<.001
One-time behaviors	Personal norms	0.382	<.001
Recurring behaviors	Personal norms	0.352	<.001

3. Instruments and Implementation

This chapter contains the full set of validated scales, including pre- (and post-) testing scales related to energy culture (i.e. norms, practices, material culture and context) as well as scales to evaluate the ease of use, engagement, satisfaction, and trust with the intervention itself. The following sections present the items, answer options, and instructions for scale calculation. The final two sections (Section 3.6 and 3.7) discuss implementation and data analysis of the scales.

3.1 Norms (Psychographics)

Motivation to Engage

Q. How much does each of the following factors affect your household energy use?

- Environmental impact [Pro-social motivation]
- Societal benefit [Pro-social motivation]
- Moral obligation [Pro-social motivation]
- Cost of the energy bill [Financial motivation]
- Convenience [Self Comfort motivation]
- Habit [Self Comfort motivation]
- Comfort [Self Comfort motivation]

Answer options: Not at all = 1; A little bit = 2; Somewhat = 3; A great deal = 4.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the value of the overall scale by taking the mean of all items. Additionally, a value can be calculated for each subscale (i.e. pro-social, financial, and self-comfort motivation) by taking the mean of all items within each subscale.

Energy Literacy

Q. Please indicate how much you agree or disagree with the following statements.

- I think about my household's energy use
- I consciously make decisions to minimize my energy use

Answer options: Almost never = 1; Rarely = 2; Sometimes = 3; Often = 4; Almost always = 5.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the overall scale by taking the mean of all items.

Energy Knowledge

Q. Please indicate whether or not you agree with the following statements.

- Turning the thermostat up higher will make the room get warmer more quickly (F)
- Ceiling insulation can help keep your home warm in winter and cool in summer (T)
- Reducing my water usage does not save energy (F)

- Energy efficient light bulbs use less than half of the electricity of incandescent bulbs (T)
- Washing clothes at lower temperatures uses less energy than higher temperatures (T)
- Leaving a window open on an upper floor lets heat escape on a hot day (F)
- The refrigerator uses more electricity than the air conditioner (F)

Answer options: Correct answer = 1, Incorrect answer/Don't know = 0.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the value of the overall scale by taking the mean of all items.

Concern and Connection

Q. Please indicate how much you agree or disagree with the following statements.

- I think of myself as part of an ecological community [Connection]
- I often feel a strong connection to nature [Connection]
- If things continue on their present course, we will soon experience a major ecological catastrophe [Concern]
- Climate change is a problem for society [Concern]

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the overall scale by taking the mean of all items. Additionally, a value can be calculated for each subscale (i.e. connection and concern) by taking the mean of all items within each subscale.

Personal Norms

Q. Please indicate how much you agree or disagree with the following statements.

- Each individual has a responsibility to do his or her part for the environment
- I don't see any problem with using a lot of energy*
- I feel morally obliged to reduce my energy use, regardless of what other people do

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the overall scale by taking the mean of all items.

Social Norms

Q. Please indicate how much you agree or disagree with the following statements.

- People in my community expect me to do my part
- My neighbors are trying to conserve energy

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the value of the overall scale by taking the mean of all items.

Performance Efficacy

Q. Please indicate how much you agree or disagree with the following statements.
- I can invest the time and effort to make changes towards reducing my energy use.
- I have the right skills to make informed decisions about how to manage my home energy use.

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the overall scale by taking the mean of all items.

Response Efficacy

Q. Please indicate how much you agree or disagree with the following statements.
- If enough people use less energy, we can benefit the natural environment
- If enough people use less energy, it will have a positive societal impact
- The amount of energy I use has an important impact on the natural environment

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

Scale calculation: Check reliability using Cronbach's alpha. If alpha is sufficiently high between the items, calculate the overall scale by taking the mean of all items.

3.2 Practices (Behaviors)

Recurring behaviors and behavior intention

Recurring behaviors

Q. How frequently do you…
- Limit time in the shower
- Cover pans while cooking
- Fill kettle only with required water
- Turn off lights when not needed
- Turn off appliances when not in use
- Unplug appliances when not in use
- Unplug rechargeables once recharged
- Reduce heating in unoccupied rooms
- Line dry laundry
- Wash laundry on colder settings
- Wait for full load before doing laundry
- Adjust thermostat setting at night or while away
- Clean or replace filters on your HVAC system

Answer options: Almost never = 1; Rarely = 2; Sometimes = 3; Often = 4; Almost always = 5.

Scale calculation: Take the sum of the items to create a total score.

RECURRING BEHAVIORAL INTENTIONS

Q. Please indicate how much you agree or disagree with the following statement.

During the next 6 months, I intend to…

- Limit time in the shower
- Cover pans while cooking
- Fill kettle only with required water
- Turn off lights when not needed
- Turn off appliances when not in use
- Unplug appliances when not in use
- Unplug rechargeable electronics once the batteries are full
- Reduce heating in unoccupied rooms
- Line dry laundry
- Wash laundry on colder settings
- Wait for full load before doing laundry
- Adjust thermostat setting at night or while away
- Clean or replace filters on your HVAC system

Answer options: Almost never = 1; Rarely = 2; Sometimes = 3; Often = 4; Almost always = 5.

Scale calculation: Take the sum of the items to create a total score.

ONE-TIME BEHAVIORS AND INTENTION

Q. Please indicate whether you have done each of the following since moving into your home

- Installed insulation (ceiling, under-floor)
- Installed high efficiency windows
- Insulated the hot water tank
- Replaced incandescent bulbs with CFLs or LEDs
- Replaced a water heater with a more energy efficient one
- Replaced refrigerator with more energy efficient one
- Replaced washing machine with more energy efficient one
- Replaced pool pump with more energy efficient one
- Replaced a pool heater with a more energy efficient one
- Replaced furnace with a more energy efficient one
- Sealed drafts around doors and windows
- Changed the air filters in HVAC unit
- Installed low-flow showerheads
- Reduced the temperature on the water heater
- Adjusted your programmable thermostat to automatically change temperature throughout the day (e.g., when no one is home, at night)
- Installed timers or sensors to control lighting
- Recycled extra refrigerator or freezer
- Replaced clothes dryer with more energy efficient one
- Installed a whole house fan
- Replaced the central AC with more energy efficient one
- Replaced your room AC with more energy efficient one
- Recycled your room AC
- Installed an evaporative cooler
- Installed a heat pump
- Installed faucet aerators
- Installed a water efficient toilet
- Replaced your range with more energy efficient one

- Installed an electronic ignition on your heating system
- Installed a heat recovery water heating system

Answer options: [Have done it; Have not done it; Planning to do; NA]

- One-time behavior scale calculation: Create a new variable for each item such that "have done it" = 1, all else = 0. Take the sum of the items to create a total score.
- One-time behavioral intention scale calculation: Create new variable for each item such that "planning to do" = 1, all else = 0. Take the sum of the items to create a total score.

3.3 Material Culture

Appliance Ownership

Q. Which of the following appliances do you own?

- Central heating system
- Central cooling system
- Room air conditioners (small units that sit in a window to cool one or more rooms)
- Water heater, gas
- Water heater, electric
- Dehumidifier
- Dishwasher
- Washing machine
- Clothes dryer
- Portable space heater
- Stove/oven, gas
- Stove/oven, electric
- Secondary refrigerator
- Separate freezer

Answer options: [Don't have; Have but don't use; Have and use]

- Have scale calculation: Create a new variable for each item such that "don't have" = 0, all else = 1. Take the sum of the items to create a total score.
- Have and use scale calculation: Create new variable for each item such that "have and use" = 1, all else = 0. Take the sum of the items to create a total score.

3.4 Context

Physical Context

Q. What type of dwelling do you live in?

Answer options: House; Apartment; Manufactured/Mobile home; Other (please specify)

Q. Do you rent or own your home?

Answer options: [Rent; Own; Other (please specify)]

Q. Were there any MAJOR changes in your home in the last 12 months that would have affected your electricity usage (e.g., a new person coming to live in your home, someone moving out, a new appliance, or an old one that was removed)? Please describe them below.

Q. About what year was your home built?

Q. What is the approximate size of your home?

Answer options: [0 to less than 500 sq ft; 500 to less than 1,000 sq ft; 1,000 to less than 1,500 sq ft; 1,500 to less than 2,000 sq ft; 2,000 to less than 2,500 sq ft; 2,500 to less than 3,000 sq ft; 3,000 to less than 3,500 sq ft; 3,500 to less than 4,000 sq ft; 4,000 sq ft or larger]

Q. How many separate rooms does your home have? Include bathrooms but exclude closets and hallways.

Answer options: [0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 or more]

Q. Do you have solar PV (photovoltaic) panels installed?

Answer options: [Yes; No; Not sure]

Q. What percentage of your household's electricity needs would you estimate your solar PV is providing, annually?

Answer options: [1 - 25%; 26 - 50%; 51 - 75%; 76 - 100%; More than 100%]

DEMOGRAPHICS

Q. What is your marital status?

Q. What is the highest level of school you have completed or the highest degree you have received?

Q. What is your annual household income?

Q. Which categories describe you? *[check all that apply]*

- White
- Black or African American
- American Indian or Alaskan Native
- Hispanic or Latino
- Asian
- Middle Eastern/North African
- Native Hawaiian or other Pacific Islander
- Other (please specify)

Q. During which time(s) of day is your home usually occupied? *[check all that apply]*

- Morning
- Midday
- Afternoon
- Evening
- Night

Q. How many people live in your home? Include yourself and people who are full-time residents of the house (e.g. spends more than 50% of their time at the house).

Q. For each person that lives in your home, please indicate their:
- Age
- Gender

Q. How many years have you lived in your home?

Q. How many years do you plan to live in your home?

Q. What is your zip code?

3.5 User Experience

Following the provision of a behavior intervention, the following questions can be used to evaluation aspects of that intervention (e.g. information, energy feedback, etc.) to determine how useful and usable the intervention is perceived to be. Note that items in italic were presented in the original instrument (Karlin & Ford, 2013a). Additional items (non italicized) have been added since then and are currently being field tested in the US, Canada, and New Zealand, but have not yet been published.

UPscale - Ease of Use

Q. Please indicate how much you agree or disagree with the following statements.

Item	Construct
I am able to get the information I need easily	*Efficiency*
I think that I can use this information quickly	Efficiency
I find this information unnecessarily complex*	Complexity
*I think this information is difficult to understand**	*Interpretation*
I feel very confident interpreting/using this information	*Interpretation*
*A person would need to learn a lot in order to use this information**	*Learnability*
I think that I would need assistance to be able to use this information*	Learnability

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

UPscale - Engagement

Q. Please indicate how much you agree or disagree with the following statements.

Item	Construct
I gained information that will benefit my life	*Relevance*
The information presented is relevant to my daily life	Relevance
*I do not find this information useful**	*Usefulness*
Most people I know would not find this information useful*	Usefulness
*I would not want to use this information**	*Intention to use*
I think that I would like to use this information frequently	*Intention to use*
I think others would like to use this information if provided to them	Intention to use

The information was provided in a fun manner	Hedonic
I find the information interesting	Interest

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5. Reverse code asterisked items.

UPSCALE - TRUST

Q. Please indicate how much you agree or disagree with the following statements.
- I trust the information
- I find the source of this information to be trustworthy

Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5.

Scale calculation: Check reliability of UPscale (i.e. all items from ease of use, engagement, trust) using Cronbach's alpha. If alpha is sufficiently high between the items calculate the value of the overall scale by taking the mean of all items.

Additionally, a value can be calculated for each subscale (i.e. ease of use, engagement, trust) by checking reliability of each subscale and, if appropriate, taking the mean of all items.

SATISFACTION

Q. Please indicate how much you agree or disagree with the following statements.
- Overall, I am satisfied with this information
- I think that most people who use this information will be satisfied with it

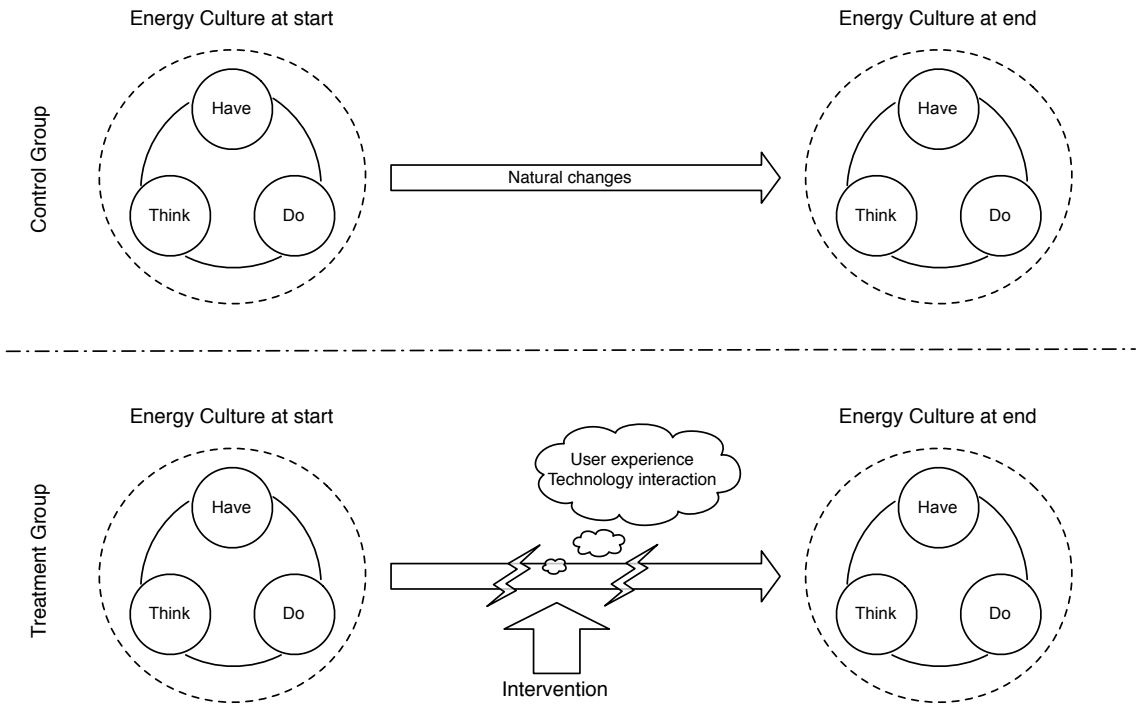
Answer options: Strongly disagree = 1; Disagree = 2; Neither agree nor disagree = 3; Agree = 4; Strongly agree = 5.

Scale calculation: Check reliability of using Cronbach's alpha. If alpha is sufficiently high between the items calculate the value of the overall scale by taking the mean of all items.

3.6 IMPLEMENTATION

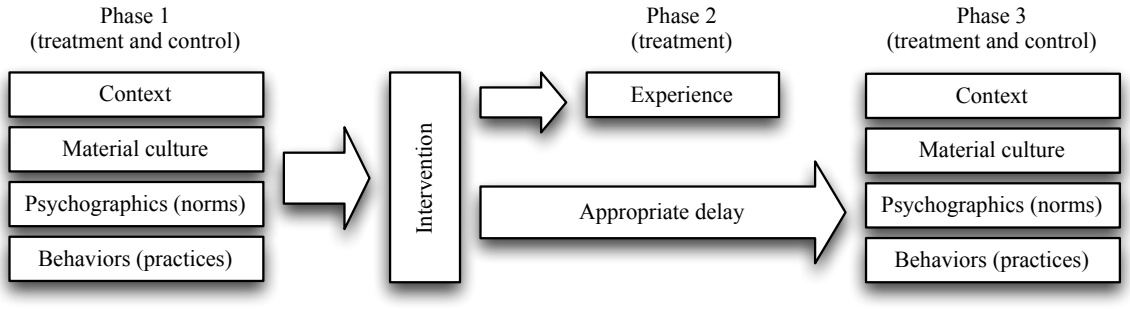
It is recommended that behavior based interventions are evaluated using a randomized controlled trial in which the change in energy culture of the treatment group is compared to that of the intervention group using pre- and post- treatment evaluation of both groups, as illustrated in Figure 4.

FIGURE 4: EVALUATION OF TREATMENT AND CONTROL GROUPS



Survey scales should be administered in three distinct phases (see Figure 5).

FIGURE 5: EVALUATION PROCESS



In the first phase, prior to the intervention being run out across the treatment group, both the treatment and control group are provided a survey containing questions relating to context, material culture, psychographics, and behaviors to: (1) describe the sample, (2) test for representativeness and subpopulations, and (3) obtain baseline measurements of energy culture against which subsequent changes can be evaluated.

In the second phase, which should occur shortly after the intervention is run out to ensure that intervention materials are still fresh in participants' minds, the treatment group is surveyed to evaluate their experience of the intervention. Because this may have an impact on overall effectiveness, it is important to capture customers' perceptions relating to ease of use, engagement, satisfaction, and trust, particularly if the intervention is implemented using a two-by-two research design, which presents opportunities for exploring differences between the treatment groups. The analysis of this is discussed further below.

In the third phase, which is designed to measure changes in household energy culture compared to the baseline measurements in phase 1, questions relating to material culture, psychographics, and behaviors are asked of both treatment and control groups. It is often not necessary to ask context questions again as these are unlikely to have changed during the intervention. For some interventions (e.g. those that target practice changes, or that aim to increase energy literacy) it may be appropriate to collapse phases 2 and 3. For other interventions (e.g. home audit programs, programs that provide a range of interventions at staged time intervals) it may be appropriate to conduct the phase 3 evaluation at additional points in time to explore how energy culture may continue to shift over time.

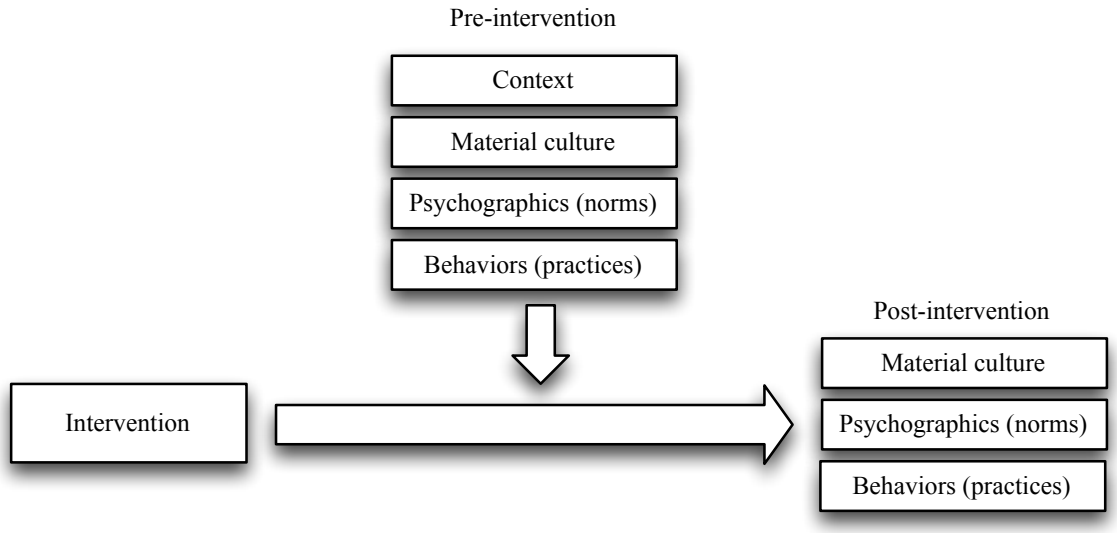
Additionally, depending on the particular intervention, it may not be necessary to include all the scales. For example, if a program is not designed in increase knowledge, and no aspect of the intervention serves to increase knowledge, the evaluators may choose not to include the energy knowledge questions. However, if knowledge was an outcome variable of interest, or if it was hypothesized to mediate the effectiveness of the intervention (i.e. the intervention led to increased knowledge, which led to a change in specific behavior) then this scale should be included in the evaluation instrument.

3.7 Analysis

Having computed values for each of the independent constructs being tested (including psychographics, material culture, and behaviors) at Time 1 and Time 2, mixed model ANOVAS and ANCOVAS can be run to assess whether the changes from Time 1 to Time differ significantly between treatment and control groups. Potential moderators (e.g. gender, renters vs. owners) can be included as fixed factors.

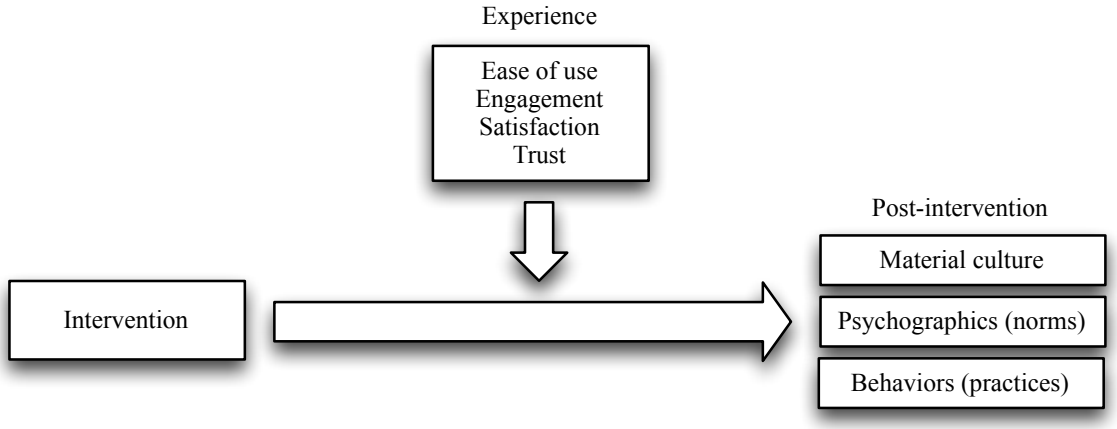
Additionally, moderation testing is suggested to assess whether pre-intervention energy culture (i.e. context, psychographics, material culture, and/or behaviors) affects post-intervention measures, as illustrated in Figure 6.

FIGURE 6: MODERATION PATHWAY



If the intervention is implementing multiple treatment groups, moderation testing can also be used to assess whether customers' perception of the intervention has an effect on intervention effectiveness, as depicted in Figure 7.

FIGURE 7: EXPERIENCE AS A MODERATOR OF EFFECTIVENESS



Finally, mediation analyses can be conducted using standard procedures (described in Baron and Kenny, 1986) to test what changes in psychographics or material culture can explain the effectiveness of the intervention. In cases of a failed intervention, such analyses can help pinpoint why a program may not have succeeded. For example, if a primary focus of an intervention was to increase energy literacy, one can test whether literacy increased as a result of the manipulation, and whether those increases predict changes in energy behavior.

4. Discussion

4.1 Importance of Understanding Behavioral Interventions

Recent legislation in California (e.g., AB32, SB X1-2) has created an environment in which investor-owned utilities (IOUs) are both mandated and incentivized to engage consumers in demand-side management. While the IOUs have successfully utilized programs such as energy efficiency rebates, direct install programs, and appliances codes to reduce energy use in California, the energy savings that can be achieved through these programs diminish as the state becomes increasingly energy efficient. As a result, behavior based programs are now being looked to as a promising strategy to capture previously untapped energy savings. Scientists and practitioners agree that significant energy savings can be achieved through behavioral interventions.

However, there is great variation in savings from the behavioral studies conducted to-date and simply collecting data on energy usage from treatment and control groups does not address the knowledge gap presented by this variation. In their review of intervention studies aimed at household energy conservation, Abrahamse et al. (2005) found that "underlying determinants of energy use and energy-related behaviors have hardly been examined".

While this situation has improved in recent years with the proliferation of privately contracted evaluation research, significant variation remains in the variables collected and specific questions used, making comparisons across studies difficult. No standard measures or metrics currently exist to conduct such assessment. Such consistency would improve our overall ability to account for variation in treatment effects and verify savings. This report aims to address this need.

4.2 Additional Items for Future Consideration

The scales presented are part of a larger "toolkit" of instruments that can be used in assessing behavioral interventions. As such, additional "tools" may be added over time and by various stakeholders in the science, utility, regulatory, and/or practitioner community. The key here is that any scales or instruments added to the "toolkit" be psychometrically tested and validated in the same way that the scales in this report have been tested. The following are two constructs that have been suggested for future testing and possible inclusion in the toolkit.

Energy Knowledge

Although knowledge questions have been developed, tested, revised, and presented above, we believe that further testing of additional knowledge questions may prove useful.

Firstly, several of the items showed restricted variability with over 80% of participants choosing the correct answer, making these items candidates for being

removed from future iterations. This highlights an opportunity for testing additional items that may have a greater sensitivity.

Secondly, the literature suggests that while general knowledge does not predict behavior, knowledge specific to the domain in question should be related to behavior in that domain. Our analysis found that knowledge was a particularly good predictor of recurring behaviors and behavioral intentions. They also significantly predicted the number of one-time behaviors respondents had already completed, suggesting that the knowledge questions included in the toolkit were relevant to behaviors of interest. However, in field testing the behaviors of interest may extend beyond those about which knowledge questions have been developed, highlighting the opportunity to develop an additional bank of knowledge questions that could be used according to the behaviors of interest. Some additional knowledge questions that warrant future testing appear below.

Please indicate whether or not you agree with the following statements.

- Electronics don't use energy once the switch is in the "off" position. (F)
- Leaving a fan on in an unoccupied room keeps the room cool. (F)
- You'll save energy when boiling water when you start with hot water from the tap. (F)
- Using a screen saver will not reduce your energy usage in the office. (T)
- Closing vents and registers in unused rooms will not result in any energy savings. (T)
- Washing dishes by hand uses more water than using a dishwasher. (F)

Answer options: [Disagree; Agree; Don't know]
T = True, F = False

Responses would be coded and evaluated in the same way as energy knowledge questions tested to date, but, to be included in further iterations of the toolkit, testing similar to that reported in Section 2 of this report is advisable.

SOCIAL DESIRABILITY

Although social desirability was not included in initial testing, follow-up conversations with SCE identified this metric as one of interest, highlighting the value of including a small number of items in the Toolkit to measure respondents' tendency to give socially desirable answers. Participants in a behavioral intervention may feel some pressure to appear more supportive of the intervention than they actually feel. A measure of social desirability would allow evaluators to control for this factor in analyses.

There are several approaches to measuring social desirability in the literature; none of them are entirely satisfactory. The most widely used measure is the Marlowe-Crowne Social Desirability Scale (a short form was developed by Reynolds, 1982). A domain-specific measure of socially desirable responding relevant to environmental issues is represented by the external subscale of the Environmental Attitudes Scale (Ebenbach, 1999). This subscale was used as the measure of socially desirable responding in the tests of convergent and divergent validity reported here. Finally, developmental psychologists have developed ways of measuring the extent to which teens give in to peer pressure. Many of the items in peer pressure scales refer directly to behaviors relevant only to a teen audience (skipping class, underage drinking). However some are more general. Items from all three sources are presented in Table 3.

Further testing will have to determine whether a more general or a more domain-specific approach (or a combination of the two) is most effective. A domain specific approach has the advantage of being more targeted and also more obviously relevant to program administrators. A general approach has the advantage of being less obvious and transparent. We have indicated items in Table 3 that we believe are particularly worthy of testing in future iterations of the Toolkit development.

TABLE 5: SOCIAL DESIRABILITY QUESTIONS

ITEM	FLAGGED FOR FUTURE TESTING
SOURCE: MARLOWE-CROWNE SHORT FORM (REYNOLDS 1982)	
It is sometimes hard for me to go on with my work if I am not encouraged. (R)	
I sometimes feel resentful when I don't get my way. (R)	
On a few occasions, I have given up doing something because I thought too little of my ability. (R)	
There have been times when I felt like rebelling against people in authority even though I knew they were right. (R)	
No matter who I'm talking to, I'm always a good listener.	*
There have been occasions when I took advantage of someone. (R)	*
I'm always willing to admit it when I make a mistake.	
I sometimes try to get even rather than forgive and forget. (R)	
I am always courteous, even to people who are disagreeable.	
I have never been irked when people expressed ideas very different from my own.	
There have been times when I was quite jealous of the good fortune of others. (R)	*
I am sometimes irritated by people who ask favors of me. (R)	
I have never deliberately said something that hurt someone's feelings.	*
SOURCE: PEER PRESSURE ITEMS (SANTOR ET AL., 2000)	
At times, I've broken rules because others have urged me to.	
I often feel pressured to do things I wouldn't normally do.	
I'd do almost anything to avoid being seen as a loser.	
I've bought things because they were the "in" thing to have.	*
I usually do what I am told.	
I rarely follow the rules. (R)	*
ENVIRONMENTAL ATTITUDES EXTERNAL SUBSCALE (EBENBACH 1999)	
I try to appear pro-environmental to please others, but I really don't believe environmental issues are important.	*
I try to act pro-environmentally because of pressure from others.	*
If I did something that might harm the environment, I would be concerned that others would be angry with me.	
I do not attempt to appear pro-environmental to others. (R)	*
It is not important for me to appear pro-environmental to others. (R)	*
I try to express only my pro-environmental views in order to avoid negative reactions from others.	
Because of today's PC (Politically Correct) standards, I try to appear pro-environment.	

(Note: Items with an (R) are reverse scored)

4.3 Conclusion

As more and more utilities and regulatory agencies focus their attention on behavioral interventions in demand side management programs, there is an urgency to ensure that evaluation of such programs is done in as rigorous a manner as possible. While the metric used to measure whether these various programs work (kWh) has been fairly standard and easy to compare between studies, the variables and metrics used to measure how and for whom they work have been left to individual researchers, with little attempts at creating a replicable model. Such standardization is common in related fields such as education and psychology, but has yet to take hold in energy program evaluation. The current report proposes a set of scales that can be used consistently with behavior-based programs including but not limited to eco-feedback, home audits, information and rebate programs, and social games.

Designed to complement rather than replace traditional measures of program effectiveness, the inclusion of such psychometrically validated scales to assess behavioral programs can yield useful insights into effective program design. This can help increase our ability to move beyond testing each individual intervention strategy for its effectiveness, to model and predict the effectiveness of future interventions based on an increased knowledge of how and for whom they work best.

Wide-scale use of a consistent and validated measure can improve and aggregate our overall knowledge about how and for whom behavioral programs work across the countless additional studies expected to be conducted in the coming years. Such knowledge is essential for behavioral programs for demand side management to take their rightful seat at the table of energy resources, such as fossil and alternative fuels.

REFERENCES

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. *Advances in experimental social psychology*.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211

Ajzen, I., Joyce, N., Sheikh, S., & Gilbert Cote, N. (2011). Knowledge and the prediction of behavior: The role of information accuracy in the Theory of Planned Behavior. *Basic and Applied Social Psychology*, 33, 101-117.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173 – 1182.

Barr, S., Gilg, A. W., & Ford, N. (2005). The household energy gap: examining the divide between habitual- and purchase-related conservation behaviors. *Energy Policy, 33*(11), 1425–1444.

Black, J. S., Stern, P. C., & Elworth, J. T. (1985). Personal and contextual influences on household energy adaptations. Journal of Applied Psychology, 70, 3–21.

Brooke, J. (1996). SUS - A quick and dirty usability scale. In: Jordan, Thomas, Weerdmeester, McClelland, Eds. *Usability evaluation in industry,* 189–194.

Cheng, M.-F. (2014). An examination of the value-belief-norm theory model in predicting pro-environmental behavior in Taiwan. *Asian Journal of Social Psychology.*

Cialdini, R., & Schultz, W. (2004). Understanding and motivating energy conservation via social norms. *William and Flora Hewlett Foundation*.

Costa, P.T. Jr. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual.* Odessa, FL: Psychological Assessment Resources.

Curtis, F. A., Simpson-Housley, P., & Drever, S. (1984). Communications on energy Household energy conservation. *Energy Policy*, *12*(4), 452-456.

del Carmen Aguilar-Luzon, M., Garcia-Martinez, J.M.A., Calvo-Salguero, A., & Salinas, J.M. (2012). Comparative Study between the Theory of Planned Behavior and the Value-Belief-Norm Model Regarding the Environment, on Spanish Housewives' Recycling Behavior. *Journal of Applied Social Psychology*, *42*, 2797-2833.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). Measuring endorsement of the new ecological paradigm: a revised NEP scale. Journal of Social Issues, 56(3), 425–442.

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

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. J*ournal of Personality and Social Psychology, 37,* 1097-1126.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35*, 790-806.

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1-14.

Gatersleben, B., Steg, L., & Vlek, C. (2002). Measurement and determinants of environmentally significant consumer behavior. Environment and Behavior, 34, 335–362

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, *35*(3), 472-482.

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

Han, H., Hsu, L.T., & Sheu, C. (2010). Application of the Theory of Planned Behavior to green hotel choice: Testing the effect of environmental friendly activities. *Tourism Management*, *31*, 325-334.

Han, H. (2015). Travelers' pro-environmental behavior in a green lodging context: Converging value-belief-norm theory and the theory of planned behavior. *Tourism Management*, *47*, 164-177.

Harland, P., Staats, H., & Wilke, H. (1999). Explaining pro environmental behavior by personal norms and the theory of planned behavior. Journal of Applied Social Psychology, 29, 2505–2528.

Kaiser, F.G., Hubner, G., & Bogner, F. (2005). Contrasting the Theory of Planned Behavior with the Value-Belief-Norm Model in Explaining Conservation Behavior. *Journal of Applied Social Psychology*, *35*, 2150-2170.

Karlin, B. (2012). Public Acceptance of Smart Meters: Integrating Psychology and Practice. *Proceedings of the 2012 Summer Study on Energy Efficiency in Buildings.* ACEEE.

Karlin, B., Davis, N., Sanguinetti, A., Gamble, K., Kirkby, D., & Stokols, D. (2012). Dimensions of conservation: Exploring differences among energy behaviors. *Environment and Behavior,* 46(4) 423–452.

Karlin, B., & Ford R. (2013a). *Beyond kWh: A New Tool for Assessing Behavior-Based Energy Interventions.* Report prepared for the International Energy Agency Demand Side Management Program (IEA-DSM) Task 24-Behaviour Change in DSM.

Karlin, B., & Ford R. (2013b). The Usability Perception Scale (UPscale): A measure for evaluating feedback displays. *A. Marcus (Ed.), Proceedings of the 2013 Human Computer Interaction (HCII) Conference.* Springer-Verlag.

Karlin, B., Ford, R., & Zinger, J., (2015). The Effects of Feedback on Energy Conservation: A Meta-Analysis. *Psychological Bulletin.*

Karlin, B., Ford, R., Wu, A., Nasser, V., & Frantz, C. (2015). *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.

Karppinen, H. (2005). Forest owners' choice of reforestation method: an application of the theory of planned behavior. *Forest Policy and Economics*, *7*, 393-409.

Kim, Y.J., Njite, D., & Hancer, M. (2013). Anticipated emotion in consumers' intentions to select eco-friendly restaurants: Augmenting the theory of planned behavior. *International Journal of Hospitality Management*, *34*, 255-262.

Kline, P. (2000). *Handbook of Psychological Testing*. Routledge, London.

Lam, S.P. (1999). Predicting Intentions to Conserve Water From the Theory of Planned Behavior, Perceived Moral Obligation, and Perceived Water Right. *Journal of Applied Social Psychology*, *29*, 1058-1071.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction, 7*, 57–78.

Long, J. E. (1993). An econometric analysis of residential expenditures on energy conservation and renewable energy sources. *Energy Economics*, 232–238.

Lopez-Mosquera, N., & Sanchez, M. (2012). Theory of Planned Behavior and the Value-Belief-Norm Theory explaining willingness to pay for a suburban park. *Journal of Environmental Management*, *113*, 251-262.

Lynch, D., & Martin, P. (2010). How energy efficiency programs influence energy use: an application of the theory of planned behavior. In *European Council for an Energy Efficient Economy Summer Study* 2037–2048.

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.

McKenzie-Mohr, D., 2011. Fostering sustainable behavior: An introduction to community-based social marketing. 3rd

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and social psychology bulletin*, *34*(7), 913-923.

Norman, P., Conner, M., & Bell, R. (1999). The Theory of Planned Behavior and Smoking Cessation. *Health Psychology*, *18*, 89-94.

Painter, J. Semenik, R., & Belk, R. (1983). Is there a generalized energy conservation ethic? A comparison of the determinants of gasoline and home heating energy conservation. *Journal of Economic Psychology*, *3*(3), 317-331.

Poortinga, W., Steg, L., Vlek, C., & Wiersma, G. (2003). Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology, 24*, 49–64.

Poortinga, W., Steg, L., & Vlek, C. (2004). Values, Environmental Concern, and Environmental Behavior: A Study into Household Energy Use. *Environment & Behavior, 36*(1), 70–93.

Ritchie, J. B. McDougall, G. H., & Claxton, J. D. (1981). Complexities of household energy consumption and conservation. *Journal of Consumer Research*, 233-242.

Rocheleau, C. A. (2013). Organ donation intentions and behaviors: application and extension of the theory of planned behavior. *Journal of Applied Social Psychology*, *43*, 201-213.

Rousseau, D.M., Sim Sitkin, R.B., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review 23*(3), 393-404.

Sahin, E. (2013). Predictors of Turkish Elementary Teacher Candidates' Energy Conservation Behaviors: An Approach on Value-Belief-Norm Theory. *International Journal of Environmental and Science Education*, *8*, 269-283

Sardianou, E. (2007). Estimating energy conservation patterns of Greek households. *Energy Policy*, *35*(7), 3778–3791. http://doi.org/10.1016/j.enpol.2007.01.020

Schultz, P. W., & Zelezny, L. C. (1999). Values as predictors of environmental attitudes: evidence for consistency across 14 countries. Journal of Environmental Psychology, 19, 255–265

Schultz, P.W. (2001). Assessing the structure of environmental concern: concern for the self, other people, and the biosphere. *Journal of Environmental Psychology*, *21,* 327-339.

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.

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

Stern, P. C. (2011). Contributions of psychology to limiting climate change. *The American psychologist, 66*(4), 303–14.

Thompson, D.W., & Hansen, E.N. (2013). Carbon Storage on Non-industrial Private Forestland: An Application of the Theory of Planned Behavior. *Small-scale Forestry*, *12*, 631-657

Thompson, E. R. (2008). Development and validation of an international English big-five mini-markers. *Personality and Individual Differences, 45*(6), 542-548.

Turaga, R., Howarth, R. & Borsuk, M. (2010). Pro-environmental behavior: rational choice meets moral motivation. *Annals of the New York Academy of Sciences*, *1185*, 211–24.

van Riper, C.J., & Kyle, G.T. (2014). Understanding the internal processes of behavioral engagement in a national park: A latent variable path analysis of the value-belief-norm theory. *Journal of Environmental Psychology*, *38*, 288-297

Verhallen, T. M. & van Raaij, W. 1981. Household behavior and the use of natural gas for home heating. *Journal of Consumer Research 8*, 253-257

Widegren, O. (1998). The new environmental paradigm and personal norms. *Environment and Behavior, 30*, 75–100.

Ziaei-Bideh, A., & Namakshenas-Jahromi, M. (2014). Explaining Effective Factors on Consumers Willingness to Pay More for Buying Green Products Based on the Value-Belief-Norm Theory. *The Open Access Journal of Resistive Economics.*