Measuring the Impacts of Behavior Change Programs: Filling in the Blanks

Lisa A. Skumatz, Skumatz Economic Research Associates, Inc.

ABSTRACT

Protocols for evaluating measure-based energy efficiency programs are well designed, documented, and generally agreed upon in the industry. However, although ACEEE and other organizations have estimated high potential energy savings from behavior modification programs, work on evaluation of behavioral programs is far from mature. This paper has two main objectives: presenting recommendations for best practices in behavior change measurement based on a review of the literature, and 2) providing, as an example, a detailed hands-on case study measuring behavior impact.

In today's energy efficiency marketplace, the number interventions that are strictly measure based is declining. While it is true that some interventions may have little-to-no interaction with behavior, most programs have some behavioral component, through use, O&M, or influences. Program managers are increasingly considering purely behavioral (e.g. social marketing) interventions – or various shades of behavioral-plus-measure interventions.

The authors discuss the state of current practices in experimental design and evaluation techniques for behavioral programs – pulling lessons from public health, social marketing, sustainability, energy and other fields – and present recommended protocols for behavioral programs developed from recent projects. We provide a number of hands-on evaluation examples from the energy literature– including recent work by the authors – and show application of the techniques in the real world. We address behavior retention, persistence, impacts compared to other approaches, and what is currently under-valued in behavior evaluation.

Introduction

For several decades, measure-based programs have served as the core of energy efficiency portfolios. More recently, behavioral programs are becoming more recognized for their potential to provide energy savings; according to ACEEE, behavioral actions and choices (with just current technologies) could save perhaps 30 percent or more energy in the US alone (Earhardt-Martinez 2009). The progress of behavioral interventions is slowed by concerns that these programs are less reliable – in terms of savings and lifetimes – than measure-based interventions. However, these concerns fail to recognize the fact that behavior has had, and continues to have, an important influence on the performance of measure-based programs in terms of both savings and lifetimes. There are two ways behavior enters into energy efficiency programs.

1) Behaviors influence measure programs and their performance. While it is true that some interventions (e.g. building envelope upgrades), may have little-to-no interaction with behavior, most programs have some behavioral component. Performance of energy efficient refrigerators (savings and lifetimes) is impacted by coil-cleaning behaviors; optimal performance of HVAC systems is only realized through proper heating and cooling settings; maintenance, and installation in locations with sufficient space. Purer

behavioral efforts include large scale social marketing campaigns with little to no focus on measures, and of course, there is a wide array of programs in between.

2) Behaviors lead to uptake of measure-based programs. The participation decision is behavioral in nature.

Those that put programs into two distinct camps –measure-based and behavioral – miss the fact that the former are influenced by behavior, and thus, they often overestimate the reliability of associated M&V (attributing all the impacts to equipment when it is often a combination of equipment and behavioral influences). They also tend to underestimate the measurability of behavioral programs / interventions.

Measure-based programs have established evaluation protocols.[1] Behavioral interventions are inherently more complex to measure and evaluate. However, behavioral evaluation has proceeded apace in fields other than energy efficiency (health, etc.), and provides lessons that can be adapted to the energy field. Furthermore, the existing protocols may need adjustment to take account of behavioral components and influences that may be important, even in measure-based programs, because of the influence of behavior on performance of the equipment (Skumatz et.al. 2010, Freeman and Skumatz 2012).

Behavioral Programs in Energy

A number of behavioral programs focused on energy use have been introduced over the years (Skumatz and Freeman 2011b). Examples include:

- isolated work in the 1970s and the organized Bonneville Power Administration Hood River initiatives in the 1980s,
- shorter shower (and other conservation) messages in California in the water crises in the 1990s that reduced energy use, Canadian initiatives in the late 1990s,
- the energy challenges from the California energy shortages in the early 2000s,
- a decade of Energy Star™ initiatives, and
- most recently, the "***setsuden***" (energy saving) momentum in Japan after Fukushima (Kakuchi 2011, New York Times 2011), among many others.

In the last half decade, significant new attention has turned to behavioral options, aided by changes in community and e-communication, and enhanced capabilities in energy-related software and hardware. Often-cited current examples include the OPower™ feedback programs, real-time pricing, thermostat feedback, smart meters, and other programs designed to affect behaviors and equipment use.

For the most part, these programs are not based on simple and traditional outreach; instead, they have tended to be more sophisticated efforts that incorporate some or all of the principles of social marketing – techniques that have been applied widely to curb teen drug use, reduce drunk driving, shame litterers, and reduce obesity. A number of the hallmarks of social marketing campaigns follow:

[1] Summarized in Skumatz et.al. 2010.

- combining traditional marketing techniques with sociological and psychological tools to influence a target behavior.
- going beyond the awareness focus of most traditional outreach by incorporating the identification of barriers and motivations, targeting a specific sector, and the use of tools such as social norms, prompts, and feedback.
- working to incorporate multiple "touches" to try to change and habitualize behavior change (Community-Based Social Marketing / CBSM is seldom a one-off effort).
- working to reach out through social networks (faith-based, neighbors, community partnerships, etc.) to make connections, aid in credibility and transfer, etc.

Various types of energy efficiency projects have incorporated social marketing tools, including: 1) real time feedback projects (using feedback, norms, prompts, messaging), 2) changes to the way utility bills are designed to incorporate comparisons to other households (uses norms, messaging, prompts), and 3) energy audit / measure installation (incentives, door to door outreach, social networks, norms, prompts). Other initiatives focus on influencing changes in how consumers behave to increase measure uptake, increasing overall program participation, changing the way consumers and customers think about and utilize energy, impacting maintenance and upkeep schedules, and helping build stronger customer relationships.

Despite the potential for all of these programs to significantly change behaviors and energy demand, the measurement and tracking of the behavior change programs lags far behind the tested M&V and evaluation methods applied to measure based programs. To some extent, this may be deserved. As we discuss below, a small minority of energy-related behavioral programs have conducted defensible or complete evaluations; however, in other respects, evaluators may be drawing too bright a line between their evaluation of "measure-based programs" and "behavioral" programs. Recall that only a small subset of measure-based initiatives is actually and truly independent of behavioral components.

Behavioral Programs – And their Evaluations - To Date

The authors conducted a detailed review of the behavior change and social marketing program literature, including published reports, white papers, journal articles, conference proceedings, and web reviews. A detailed summary of this work (and the associated bibliography) is provided elsewhere (Skumatz et. al. 2009, Skumatz and Freeman, 2011a). We examined the broad topic of "behavioral" programs – including work in fields beyond energy. We were interested in examining the techniques used, effective approaches, costs, and other topics. We found the greatest number of examples in the health, transportation, and recycling fields (respectively, 27%, 21%, and 18% of the summaries). However, about one-sixth of the studies we found were related to energy efficiency and behaviors, and these represented some of the best-documented of the social marketing research because energy allows for easier measurement and reporting of impacts (energy compared to reductions in teen binge drinking, for example). The figure below displays the major groups of programs we reviewed. Note also that 80% of the programs reviewed were focused on residential consumers, rather than the commercial sector.

The case studies included energy audits, mass marketing campaigns coupled with door to door efforts, residential assistance programs, feedback programs, efforts at schools and universities, public events, and public spaces, and social media campaigns.[2] Some were

[2] The case studies are reviewed in more detail (and citations provided) in Skumatz and Freeman 2012, and Green and Skumatz 2000. Space does not permit the review here.

behavioral programs alone with no measures installations (a limited number of programs) and others were efficiency efforts in which the behavioral modification portion was one component of a larger program incorporating social marketing tools. The review found impacts from a range from a low of 0% reduction in electric or gas use (and in one case an increase of residential energy demand!) to a high estimate of 30% reduction in residential energy consumption (a multi-resource audit in Ontario using incentives, coupons, in-home demonstrations, and other tools). One commercial example reported a high of slightly over 30% reduction from a program. The average energy savings in the case studies was in the 5 - 15 % range, where reported. Similar results, ranging from 4%-12% savings, were found for the case of residential real-time pricing (feedback) pilots (Foster, Mazur-Stommen, 2012). However, the case studies tend to highlight 'success stories' and finding published information on typical program impacts is extremely challenging.

Not all behavior modification programs reported progress toward goals in reduction in use or demand. Some used metrics such as units sold or distributed (CFLs, ENERGY STAR™-rated appliances), commitments or following through on commitments to behavioral modifications (turning off lights, using power strips, cold laundry, shorter showers, etc.), and still others measured success based on the number of audits completed. The costs of the programs varied significantly from smaller-scale programs less than $100K to a multi-year large scale mass media campaign with a starting budget of more than $23 million.

Only three of the studies available addressed retention of educational messages and installation of low-cost energy-efficiency measures delivered through energy education programs. The Energy Smart Program conducted in Oregon with low-income households, found strong to mild retention (about 40% after 3 years) of behavioral changes. Another early study found 85%-90% of the savings from the education portion of a weatherization program was retained after three years.

Important Gaps in the Evaluation of Behavioral and Social Marketing Programs

Our review of the literature was illuminating not only in what is included in the published research,[3] but also what has not been reported. Certainly, we identified a number of common success elements in achieving behavior change; however, overall, the evaluations were weak, omitting even the most basic principles of evaluation work in impact analysis. There were also two important evaluation topics lacking in the published reports – cost and cost-effectiveness, and retention - gaps identified in our previous review of the literature ten years earlier (Green and Skumatz, 2000). If behavioral programs are to be considered seriously in the resource mix, these represent critical gaps in the literature. Computation of behavioral savings and unit costs comparable to those available for measure- and generation-based resources is not possible without these data. Our review found:

- *Weak impact evaluation information*: Although pre-post comparisons were provided for a number of the studies, control groups or other methods addressing dynamic baseline issues were included in only a few studies. This, along with the very small sample sizes for the majority of studies, significantly weakens the credibility and transferability of the results.
- *No cost-effectiveness information*: Quite a few reports include estimates of impacts, and a few report total budgets, but this information rarely overlaps. They omit analysis of impacts per dollar spent (for energy or other impacts). In some cases this may be due to

[3] Best practices in social marketing and outreach programs were summarized in Skumatz and Freeman 2011.

confounding factors (the behavior change was part of a wide portfolio of programs) while in others it is due to a lack of a control group, baseline measurements, or clear accounting of the costs of the outreach effort.

- *No Retention Results*: An even more glaring omission in behavioral efforts is the issue of behavior retention. Estimates of measure-based useful lifetimes have been incorporated into evaluation protocols for decades. However, there is extremely limited amount of information on how long a behavioral change lasts. Even well-funded multi-year statewide outreach programs have not examined the persistence of behavior change. The implications of the retention on determining the overall cost effectiveness of the efforts are significant. Unfortunately, even if first year annual savings estimates are available, it is not possible to develop reliable estimates of the benefit-cost ratio, nor is it possible to rely on long-term savings from programs that are not continually refreshed. For this reason, many utilities assign retention values no higher than three years in most cases and a one year 'deemed' value for behavior retention is not uncommon.[4] Many programs lacked even serious attempts at evaluation; instead, they described the program, told of their message and described their attractive posters, or the like. Many of the programs seemed to suggest they hadn't evaluated the work because it 'was a pilot' or similar rationale. Without evaluation work – even of pilots – frankly, it is no wonder behavioral and social marketing programs have not been taken more seriously in resource planning.

Evaluation Methods and Principles for Behavioral Programs

While not yet widespread, additional growth of behavioral initiatives may still be hampered by perceived weaknesses in evaluation techniques. However, proper evaluation techniques exist, and have been explored and developed, including substantial work in other fields (especially health).

Measurement protocols for energy-related behavioral programs follow the same principles as many other types of behavior-related evaluation work (Sebold et.al. 2001, GAO, 2009, Skumatz et.al. 2010, Sergici and Faruqui, 2011, Skumatz 2012) and are not mysterious; the bottom line is that there are few substitutes for good up-front experimental design and random assignment.[5] An abbreviated summary of best practices principles follows. Although phrased in terms of residential programs, the principles extend to other programs and sectors.

Identifying goals and conditions. Identify the goals of the program and the effects of interest (including a definition of what constitutes "participation" or "adoption"), and ensure that the effects can be seen to be caused by the program's intervention(s) (and not spurious factors). Assure that the program is administered to a group of participants that can be seen to represent the (ultimate) population of interest.

Experimental design and sampling. Plan for a test and control group. Both the control and test groups should be large enough to support statistically valid and meaningful comparisons. Sample sizes supporting +/-5-10% at 90-95% confidence are preferred.[6] The control group

[4] Although it is unclear if a median EUL of 3 years for behavior modification can be justified or whether the 1 year assumption undervalues the impacts given that there is minimal research for this estimate.

[5] This is the approach illustrated in the case study conducted by the authors that is described later in this paper.

[6] Assuming large populations, these requirements tend to require sample sizes of 68, 96, 270, or 380 observations, with higher numbers preferred. Greater specificity can be provided on sample size needs depending on the degree to which the measurement needs to address Type 1 error, Type 2 error, one or two sided hypothesis testing, single or repeated measures experiments, etc. However, a surprising number of social marketing programs have measured the

should be as similar as possible to the test subjects (and both groups should be as similar as possible to the ultimate group that will be eligible for the program to maximize transferability of results). Pre-post measurement of the test group is not best practice. Pre-post alone is vulnerable to seasonal differences, and other factors; control groups allow easy and reliable netting out of these variations. The control group sorts out impacts from effects beyond the program (e.g. nationwide ads from EPA or others, etc.)– and serves as a dynamic baseline against which the effects can be measured to provide net impacts. The premier experimental design is random assignment of eligible customers into the test and control groups (Sebold, et.al. 2001). Random assignment also helps to eliminate self-selection bias. Other approaches that have been taken include use of "similar" counties or cities, neighboring / similar states, etc.[7] Controlling for other factors from these "similar" control groups can be attempted through corrections with statistical models, but random assignment is much more straightforward and reliable.

Measurement design. Evaluation methods need to be clearly laid out before any data collection is conducted. When evaluation is concluded, all limitations of methods and results need to be clearly identified. In addition, the evaluation should include an assessment of the associated uncertainty. Identify the way in which the impact(s) and costs will be measured. For energy behaviors and energy savings, there are several main approaches:

- *Metering*: If the project (and budget) allows, metering the equipment affected by the desired behaviors over the course of the experiment provide direct and reliable information on the behavior change and its energy impacts. With large budgets, metering may be installed in large[8], random (or representative) samples of the test and control groups; with more modest budgets, metering samples are small, strategic samples that can be generalized to larger samples.
- *Utility bills and impact evaluation:* Preferred data for this option includes monthly energy usage (and billing cycles/meter reading dates and possibly tariffs) for all treatment and control customers, or alternatively, for a significant and random sample of each.
- *Surveys and reported behaviors:* The researcher needs to identify the relative appropriateness of phone, in-person, web, or other types of surveys. Research should include well-crafted / tested question methods, for example, asking about behaviors undertaken in specific time frames, rather than "general" habits,[9] and other preferred survey approaches.[10] Again, control groups are highly recommended to provide "baseline" behaviors.
- *Demographic Information:* Gathering information on number of occupants, socio-demographics, appliance data, occupancy (move-ins, etc.), weather data, and other information can help in developing statistical models that control for these sources of variations in results when conducting impact evaluation work or other comparisons.

---

impacts based on sample sizes substantially less than 100 (often 30), and only pre-post and not control group measurements, to the detriment of confidence in the results.

[7] As widespread education campaigns affecting both target and non-target audiences become more common, finding a baseline to measure against is more difficult - it is hard to uncover a population with a "zero" behavior baseline.

[8] Obviously the principle is that large samples reduce the variance and help detect significant differences between the groups.

[9] For example, 'did you use the power strip yesterday', or 'how many of your last two laundry loads used cold water', rather than 'do you use power strips' or 'do you use cold water for your laundry'.

[10] To help increase confidence in the survey reports, the researcher might also research the "say / do" gap and run scenarios on the range, conduct a sample of on-sites to confirm some behaviors that might be easily observed (current laundry temperature settings, etc.), and benchmark against a sample of billing data, where possible.

Impact analysis. The basic preferred analysis approach is a comparison of means between treatment and control group, using either one pre/post period, or periodic measurements. The appropriate tests for statistically significant differences are performed to identify impacts from the program. Multiple measurements over the course of the project / pilot provides advantages in efficiency and variance reduction (due to the correlation between measurements at different time points), and thus, greater confidence in the results. Analysis of the data up-front is valuable (monthly comparisons, plotting data, and conducting comparisons of "features" between test and control groups to ensure comparability). Impact evaluation work using statistical models and the energy data can provide reliable estimates of these means. This may employ one of several methods for estimating impacts:

- *Measurement and Verification (M&V):* using metering or estimating key parameters from a random sample (or all) of the participants and control group and applying to all members of the group.
- *Statistical Analyses:* applying statistical regression models[11] to utility billing[12] or metering data of all program participants, including approaches like differences of means / ANOVA, difference in differences[13] and panel data regression analysis[14], and other methods provide reliable estimates of impacts. Cross section and time series approaches are valid. There is an extensive literature on statistical, or statistical / engineering adjusted models.
- *Surveys and Self-Reporting*[15]: Surveying certain populations to gather information regarding knowledge or behavior to estimate the savings-related changes from behavioral / educational / social marketing programs, and analyzing for statistical differences in the adoption of the behavior. Assuming energy savings and energy are the key impacts of interest, an additional step identifying an estimated or deemed value for the savings "per adopted behavior" may be the best information on overall savings available from this method.

[11] Econometric texts can address the issues involved in regression modeling including model misspecification, measurement error, correlations, etc.

[12] Billing analysis using weather-normalized consumption data provided by the utility commonly is used to estimate gross savings. Billing analysis requires consistent residency for two or more years, so one year of pre-program data can be compared with one year of post-program data. Billing analysis may be used to estimate gross savings of education programs combining low-cost measures and behavior modification. However, as billing data are inherently too "noisy," gross savings less than 10% of pre-consumption levels are hard to detect. (Skumatz et.al. 2010)

[13] This involves netting out mean differences between treatment and control groups in the pre-treatment period from the mean differences between treatment and control groups in the post treatment period. If the difference in differences of the mean values is statistically significant, then the treatment is found to yield an observable effect in the usages of the treatment customers. (Sergici and Faruqui, 2011)

[14] Advantages of this approach include the possibility to increase the efficiency and precision of the estimate using repeated measures on each program participant and to account for time invariant unobservable variables that would otherwise lead to biased estimates. Modeling approaches include fixed effects or random effects models (Sergici and Faruqui, 2011)

[15] Self reported data are often augmented with site visits and selected metering (e.g., hours of use).There are many texts that address the issues in survey design, question development, bias reduction, etc. This is not addressed here.

- *Costs and Other Effects:* Note that we do not consider "impacts" in terms of energy or similar all the analysis that is needed. These evaluation efforts need to include strong cost tracking to support analysis of cost per impact, cost-effectiveness, and benefit-cost type-calculations that will support analysis of the impacts in context and in comparison to other programs and strategies for delivering energy and energy efficiency. This should be built in from the beginning, as should any relevant measurements needed for important non-energy benefits, if relevant.

Retention analysis. Measurement of savings from behavioral programs over the course of a full year for behavioral programs is preferred, to help account for seasonal effects. However, we would recommend working to put in place a measurement protocol that follows beyond that period to test for retention of the effects – which is a very important special uncertainty component of behavioral programs.[16] The retention of an adopted behavior change lasts only as long as the behavior remains changed,but the impact on savings may be suffer because of total cessation of the behavior, occasional retention by an actor, or retention by only some occupants of the home. Conversely, some behaviors may form new habits and remain in place for a lifetime. All of these possible changes – and more – will have an effect on the lifetime (and level of) of the estimated savings from the program.

Considerations and alternatives.[17] Although random assignment is the "gold standard", the world (and budget) does not always allow for this design – particularly if large-scale broadcast media are used, and potential participants cannot feasibly be excluded. Other options include:

- *Quasi-experimental comparison groups,*[18]
- *Statistical analysis of observational data*,[19] or
- *In-depth case studies or other approaches.*

Applying Behavioral/Social Marketing Evaluation – An Example

In 2009 and 2010 the authors conducted a social marketing project covering 1,600 households in suburban Colorado (Skumatz and Freeman 2011b). The project was tasked with delivering conservation – in the form of energy and recycling. However, as part of the project, we built in experimental design that would support detailed and defensible evaluation of the results – including the two gap areas of cost-effectiveness and retention.

[16] The impact of behavioral and other influences on the retention for measure-based programs, or the combined measure and behavioral programs, is being recognized and addressed in some cases and by some agencies (Skumatz 2012).

[17] Note that the GAO report also stated that improvements to any evaluation can be achieved by: collecting additional data, targeting comparisons, and gathering a diverse body of evidence (GAO 2009).

[18] These resemble randomized experimental design, but the groups are selected from un-served members. This might include groups denied participation when a program is full or those that will participate in the next period, etc. According to the GAO report (GAO, 2009), the approach requires statistical analysis to establish groups' equivalence at baseline, and potentially, specialized statistical modeling in some examples, regression discontinuity analysis (GAO 2009)

[19] This approach requires observing and collecting data prior to the intervention, and after the intervention

Identifying Goals and Conditions

Program goals were to lead to the adoption of energy efficiency and recycling behaviors by households in the neighborhood through outreach and social marketing, not program-provided measures. Levels of participation included commitments and actual adoption of behavior change that saved energy or landfilled tons[20] or avoided greenhouse gas emissions, with actual adoption as the primary goal. We conducted focus groups to identify impeding barriers that could be addressed, and specific behavior recommendations that could be adopted that could feasibly be expected to lead to the desired changes (identified in the tables of results below). The use of a control group was included to address the question of impacts possibly caused by "spurious" factors, and the households selected were typical of medium to higher income neighborhoods in a suburban area of Colorado's Front Range.

Experimental design and sampling

The experimental design is presented in Figure 2. The three neighborhoods, each with more than 500 households, were randomly assigned to one of three levels of outreach and social marketing interventions:

- Group 1 was our "control group", receiving only standard (limited) outreach including city-wide information on energy efficiency and an annual recycling information mailer;
- Group 2 was our "basic social marketing" group, including prompts commitments, norms, barrier identification on energy efficiency and recycling, delivered through door hangers and mail; and
- Group 3 was the "full treatment / enhanced social marketing" neighborhood, whose homes received the same social marketing outreach as Group 2, plus delivery of materials via a door to door outreach campaign and phone calls.

Figure 2. Experimental Design Summary

Three groups of comparable[21] single family households were each given different experimental treatments, with the neighborhoods assigned to the treatments randomly. Baselines were established for each Group and weekly measurements were recorded to measure impacts of the outreach on recycling and waste diversion. Energy efficiency behaviors were measured through pre and post surveys, commitments, and focus groups. There were approximately 500 households in each group for a total of 1,600 households.	
Group 1: Control or no-treatment route	Minimal level (standard) outreach materials
Group 2: Low treatment route: Social marketing-outreach and door-hangers	Received social marketing treatment and materials with the exception of door-to-door personalized visits
Group 3: Door-to-door group - Social marketing outreach materials plus door to door visits	Received the same outreach as Route 2 with the addition of door-to-door personalized visits and phone call reminders

Measurement Design

We conducted extensive baseline, on-going, and post-measurement of a variety of indicators designed to allow measurement of effects, cause, and costs. The plan is illustrated in Figure 3. Baseline measurement was necessary to measure incremental impacts attributable to

[20] A specific goal was intended; increasing recycling of 7 pounds per household per week in the test neighborhood; the goals for energy were less quantitative because energy bills were not available from the utility. Instead, significant increases in the adoption of specific energy behaviors were the goal.

[21] Households in all groups had similar demographics home size and trash and utility services.

the project's interventions, and these impacts were essential to computing the relative costs-per-impact for Groups 2 and 3 relative to Group 1 (the control neighborhood).

- *Energy Baseline:* Despite requests to the utility, we were unable to gather metered data on energy use because the state, and not the local utility, was the project sponsor. We gathered baseline information on energy equipment and behaviors from surveys, focus groups, and one-on-one household visits. The authors used surveys to conduct pre/post comparisons on metrics such as the number of CFLs in each house, the frequency / likelihood that respondents undertook energy efficient behaviors such as purchases of ENERGY STAR® appliances, reduced automobile idling time, low-flow shower head installation or reduced shower times, and others. We also gathered information on energy knowledge/awareness topics, which we compared between Groups and as pre/post comparisons.
- *Trash and recycling baseline:* Recycling was measured through baseline set outs of trash and recycling, tonnage data, calculated diversion rates, all monitored weekly before, during and after the social marketing campaign for each Group for comparison.

Figure 3. Data Collection / Measurements for Each of 3 Groups
(Group 1=control; 2=partial treatment; 3=full-treatment social marketing)

Action	Description	Pre	During	Post	Post-Post[22]
Baselines & Impacts - Energy Actions	Pre- and post surveys measuring knowledge, awareness, attitudes/beliefs, behavior occurrence (KABB); includes purchase of CFL bulbs, weatherization, use of cold water wash, powerstrips, car idling, and other behaviors, attitudes, occupancy, and other information. EE Actions taken were measured using interviews and surveys; post survey for retention.	All (1& 2&3)		All	All
Baselines & Impacts - Recycling Actions	Pre-and post surveys for actions and knowledge (KABB); weekly monitoring of tons and participation across homes, as well as periodic monitoring of individual trash / recycling tonnages and remaining potential. Baseline included one year of tonnage "before" for all three neighborhoods. 12 months of post-outreach measurement to examine retention.	All	All	All	All
Web-site visits, joining "club"	Tracked hits & other data to measure interest in the project. Monitored number of households requesting to join the conservation / recycling club, Facebook™ sign-ups by group, etc.		Group 2 & Group 3		
Commit-ments	Written commitments were collected either through the web site, or through return-mail postcards (Group 2), or collected during one-on-one visits or their returned postcards (Group 3).		2 & 3		
Cost	Cost for each round of interventions throughout the project		2 & 3		

Impacts Analysis

The analysis approach examined differences in means between the two test groups, net of the change in the control group. Given that the control group was local, randomly selected, and completely comparable, no additional sophisticated modeling was required. Impacts on campaign recall and commitments were measured. On all metrics (commitments, MTCE,

[22] For retention

recycling tons, energy behaviors undertaken), the impacts in the door-to-door neighborhood (Group 3) were markedly higher than results in the neighborhood without personal contact (Group 2). The key results on energy behaviors are shown in Figure 4. The results show between 30% and 80% increases in key energy conservation behaviors compared to the control group, and significantly higher uptakes in behaviors for the full treatment Group 3 compared to the partial treatment Group 2 (three times as large for caulking). The recycling results show that Group 2 increased recycling twice as much the control group, and Group 3 (which included the door-to-door component) increased recycling four times the increase seen in the control group (Group 1).[23] The combined actions result in more than 100 total MTCO2e of greenhouse gas per year diverted from the behaviors.

Figure 4. Energy Behaviors Reported in Follow-up Surveys

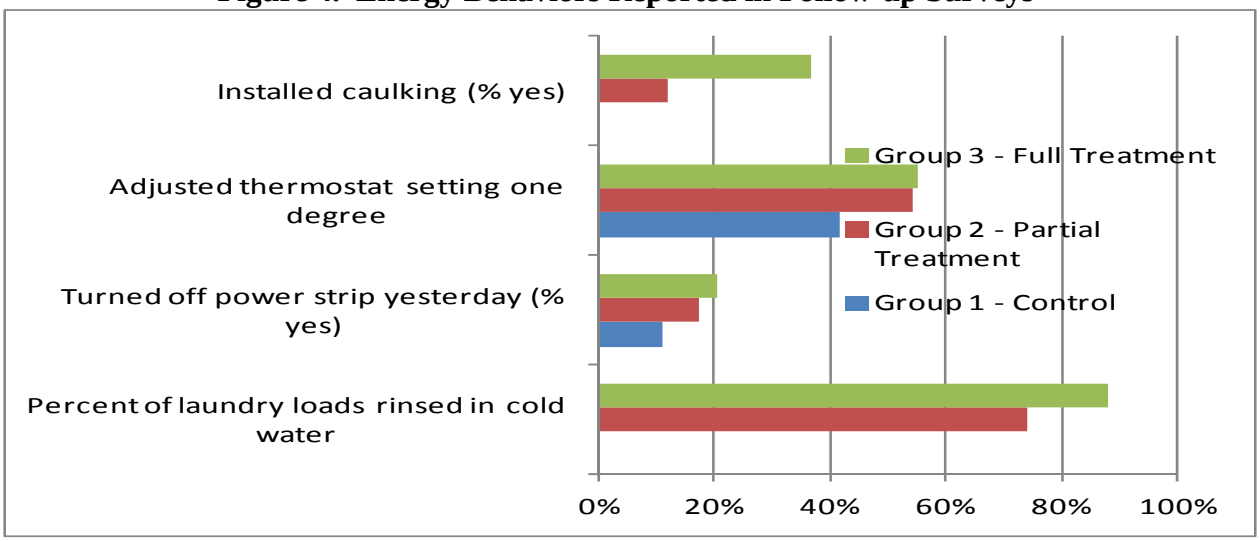
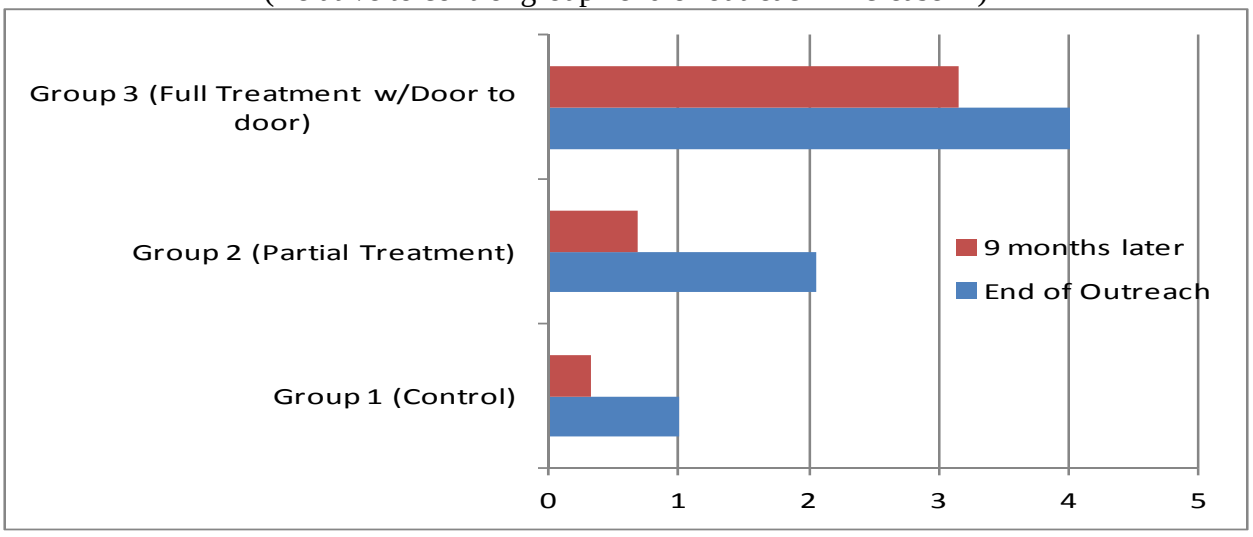


Figure 5. Recycling and Retention Results by Group – Relative to Control Group
(Relative to control group "end of outreach" increase=1)



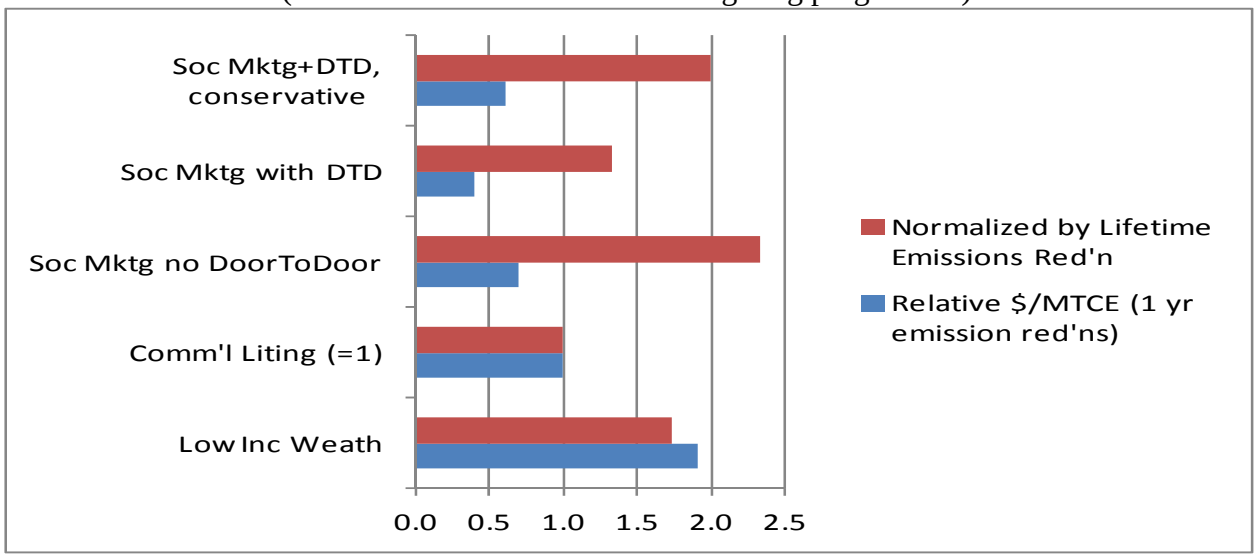
[23] It is uncertain why the control group had an increase in recycling during the outreach period. Our efforts were strictly kept from this group; however, many national organizations have education on recycling that was not under our control. Hence the need for a control group.

The costs were also analyzed. As expected, costs per household were multiple times the cost for the door-to-door work, but the higher impacts resulting from those interventions reduced the cost-per-impact to less than those in the "partial" outreach routes for each type of impact (commitments, MTCE, recycling tons, energy behaviors). However, these comparisons alone don't put the cost information in perspective. In several previous papers, the authors have computed the cost per metric ton of carbon equivalent (MTCE) from several standard measure-based programs (Skumatz 2010, Skumatz 2009). Our computations (Figure 6) from this pilot program indicate that the costs per MTCE (or relative cost per MTCO2e, since we are considering ratios) for the social marketing program would be perhaps half the cost of the cheapest measure programs – commercial lighting retrofits (Freeman and Skumatz 2012).[24] However, when program lifetime estimates are incorporated into the equations, the relative cost per MTCE for social marketing is about twice the cost of commercial lighting programs, and in the same range as low income weatherization efforts. These estimates would be revised downward for full-scale social marketing programs, and upward if the social marketing measures lasted a shorter time. We find that the cost per MTCE and potential for these types of social marketing programs – even relatively expensive small-scale pilot versions – are not unfavorable relative to traditional measure-based options.

Retention Analysis

Although we have not completed our follow-up of retention of energy behaviors, we did gather 9-month follow up data on recycling (Figure 5). We found Group 2 retained only 30% of its recycling increase, while Group 3 retained nearly 80% of the recycling increase. Surveys are currently being conducted that follow up the energy and recycling behaviors one year after the interventions.

Figure 6. Relative Costs per Impact by Group
(Relative to cost for commercial lighting program =1)



[24] Assuming a 10 year life for the commercial lighting and 3 years for the behavioral measures. Changes in these assumptions would lead to revisions in these illustrative computations.

Conclusion

Behavioral programs have the potential to deliver significant savings. ACEEE estimates 30%; most of the pilot studies have shown that even small scale efforts routinely deliver 5-15% reductions in energy use. There are legitimate concerns about behavioral programs – and social marketing – efforts.

- The savings have been well-measured (including control groups) in only a few cases, and programs are all distinct, potentially leading to different savings values. Attribution to specific program interventions is more complex than measure-based interventions.
- The programs have mostly been pilot in nature; full-scale implementation results may lead to different savings – and cost-- results.
- Costs are rarely measured, and neither is retention, hampering computation of cost per unit savings.

However, behavioral programs have several major advantages when compared with traditional measure-based programs:

- They can have significant impacts on energy use (individual pilots and social marketing programs commonly show impacts on the order of 5-15% savings) – which reflects an enormous potential realized by few purely measure-based programs. ACEEE's estimate is 30% potential lost savings from actions.
- They can be implemented quickly --with widespread adoption in a matter of weeks to months.
- They do not require programmatic purchases, delivery, or installation of equipment, intrusions into homes, and other efforts.
- The retention from social marketing is still a question. However, results indicate door-to-door methods have much stronger retention than mail-type outreach of the same materials, making the results potentially stronger than those already known for traditional outreach programs.
- Results from this project indicate the cost effectiveness may be on the order of other measure-based programs.

Integrated plans need data on energy, cost, and years the resource will be in play. Behavioral and social marketing programs have generally been small parts of these plans, at best, and may be significantly undervalued in portfolios if indicative research bears out. The two issues we identified in this and previous studies (Green and Skumatz 2000) remain cost-effectiveness and retention. More research on these questions is essential if behavioral programs and social marketing / outreach programs are to be a more integrated and reliable part of the energy portfolio, but design and evaluation techniques to address these issues are available and tested. Behavioral programs have tremendous potential, our research on relative costs show results in the range of other measure programs, and retention from at least some types of program are promising.

However, in the end, program managers need to become comfortable with behavior-based programs. There has been an artificial dichotomy between "us" (hard, reliable, measure-based programs with engineering underpinnings) vs. "them" (soft, fuzzy, unreliable, behavioral feel-good programs). The truth of the matter is that the choice between them has been a false choice. Elements of behavior are critical determinants in real-world program performance and

explain differences in carefully-measured savings reported for measure-based programs. The "measure" savings we've all been counting on are delivered from both measures and behaviors; even greater energy savings potential is possible if we embrace the full spectrum of program designs – from pure measure to pure behavioral and everything in-between, and get busy and apply good evaluation to the behavioral components. We all win if "setsuden", Japan's post-Tsunami-crisis ethic of energy savings, becomes part of our everyday lexicon as well.

References

Earhardt-Martinez, Karen. 2009. *Testimony before the United States House Committee on Science and Technology*, Subcommittee on Energy and Environment. April 28, 2009.

Foster, Mazur-Stommen. 2012. *Results from Recent Real-Time Feedback Studies.* February, 2012. ACEEE Report Number B122

Freeman, Juri and Lisa A. Skumatz, 2012. "Widgets versus Actions: Measuring the Role of Behavior Change in DSM Programs", Proceedings of the IEPEC Conference, Rome, June.

Green, John and Lisa A. Skumatz, Ph.D., 2000. "Evaluating the Impacts of Education/Outreach Programs – Lessons on Impacts, Methods, and Optimal Education". Proceedings from the ACEEE Summer Study on Buildings, Asilomar, CA, August.

Kakuchi, Suvendrini, 2011. "Energy-Saving 'Setsuden' campaign sweeps Japan after Fukusima", Guardian Environmental Network, guardian.co.uk, August 22.

New York Times, 2011. "In Japan, the Summer of Setsuden", Editorial, September 25.

Sebold, Fred, Alan Fields, Lisa A. Skumatz, Shel Feldman, Miriam Goldberg, Ken Keating, Jane Peters. 2001. *A Framework for Planning and Assessing Publically Funded Energy Efficiency*. 2001. Study ID PG&E-SW040

Sergici, Sanem, Ph.D. and Ahmad Faruqui, Ph.D., 2011,"Measurement and Verification Principles for Behavior-Based Efficiency Programs", The Brattle Group, Prepared for Opower, May.

Skumatz, Lisa A. 2009. *Do Energy Efficiency Strategies Outperform Recycling in GHG Mitigation and Job Creation?* International Energy Program Evaluation Conference. Portland, OR 2009.

Skumatz, Lisa A. 2010. *A kWh is not just a kWh: Comparing Energy Efficiency Programs in Terms of GHG, Job Impacts, and Policy Achievements*. Proceedings of the American Council on Energy Efficiency Summer Study on Buildings. Asilomar, CA. 2010.

Skumatz, Lisa A., 2012. *What Makes a Good EUL? New Protocols for Estimating Useful Lifetimes and an Analysis of Existing Estimates, Proceedings of the IEPEC Conference, Rome, June.*

Skumatz, Lisa A., and Juri Freeman, 2011a, *Getting the Most from Colorado's Recycling Programs and Infrastucture: Social Marketing Outreach and Education Toolkit – A Guidebook for Communities*, Skumatz Economic Research Associates, Superior, Colorado, May.

Skumatz, Lisa A., and Juri Freeman, 2011b, *Getting the Most from Colorado's Recycling Programs and Infrastucture: Best Management Practices, Impacts, and Cost-Effectiveness of Social Marketing*, Skumatz Economic Research Associates, Superior, Colorado, May.

Skumatz, Lisa A., M. Sami Khawaja, Jane Colby. 2010. *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence on Energy Efficiency Behavior*. Prepared for California Institute for Energy and Environment. November 2010.

Skumatz, Lisa A., Carol Mulholland, and Sharyn Barata, 2009, "Social Marketing: Evaluating the Social Twist on Marketing Programs", AESP Strategies Article, Association of Energy Services Professionals, Spring 2009United States Government Accountability Office. 2009. *Program Evaluation: A Variety of Rigorous Methods Can Help Identify Effective Interventions*. November 2009