

Behavioral Program Evaluation

Dan Violette
Jane Hummer
Ryan Firestone



Table of Contents

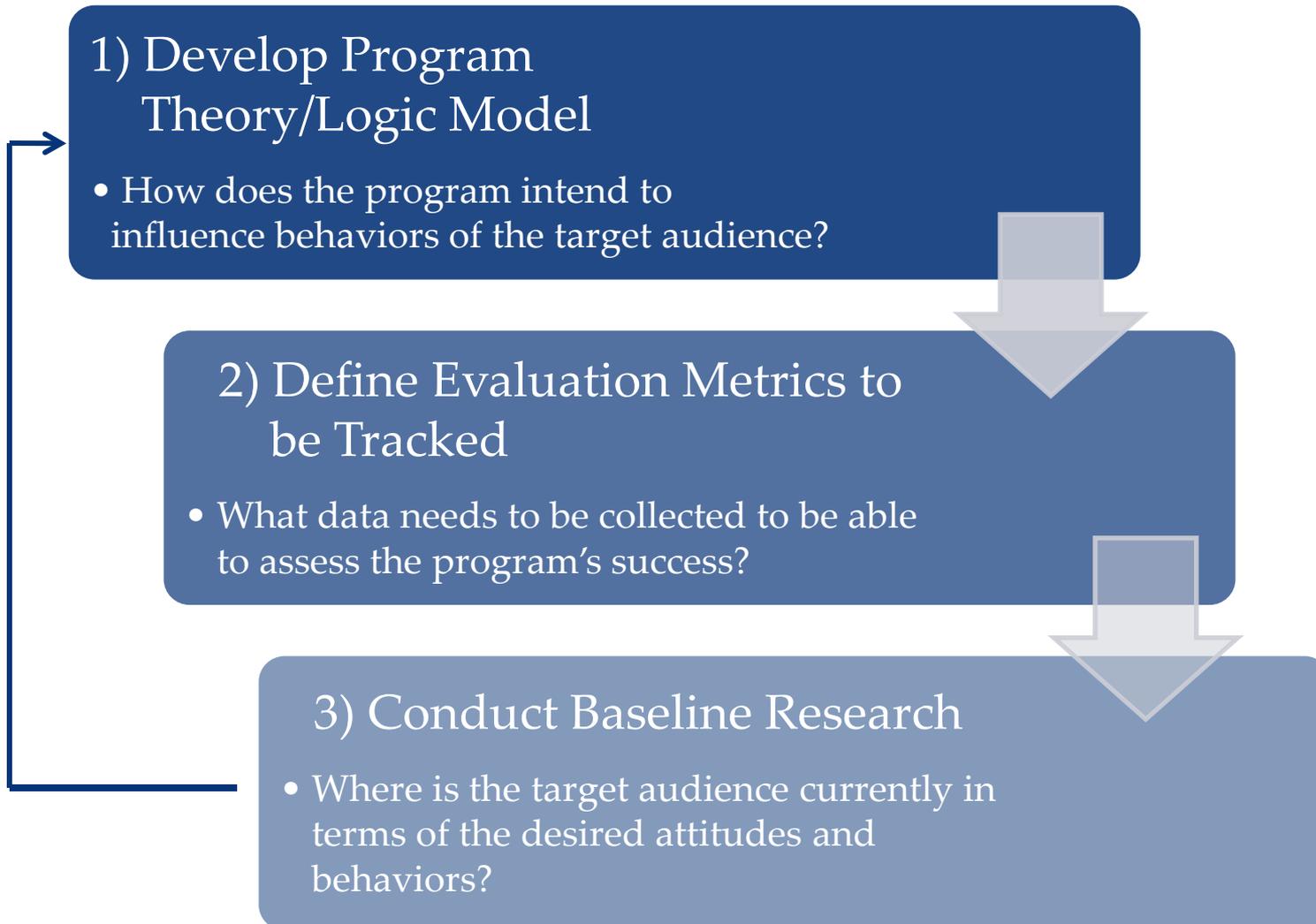
- 1 » Evaluation Planning**
- 2 » Program Activity Review**
- 3 » Outcomes and Metrics**
- 4 » Impact Evaluation**

Context: why evaluation of behavioral programs matters

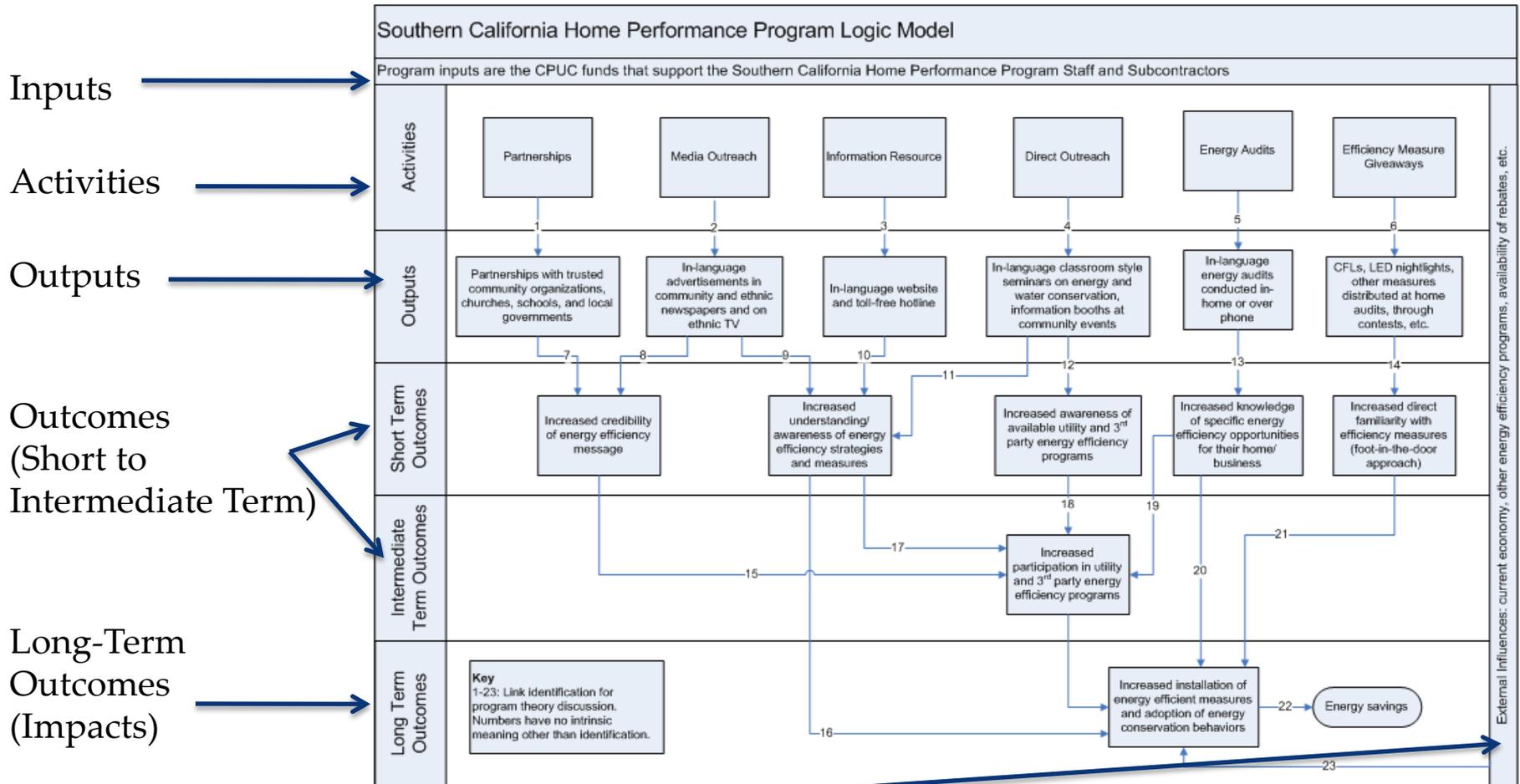
- Broad definition of behavioral programs: programs that promote voluntary changes in habitual and purchasing behaviors
- Behavioral programs are gaining momentum as we accept the fact that the strategies of the past 30+ years of energy efficiency programs have failed, particularly in the residential sector
- Reluctance to deploy these programs due to lack of solid evaluation research
- Programs are more challenging to evaluate, but not impossible; require a systematic approach to evaluation
- Program administrators need to know if they have “moved the needle” and how they can refine future iterations of the program design.
- Without this cycle of evaluation and program design refinement, taxpayers are essentially paying the program administrator to gamble on their dime.

Evaluation Planning:
Developing Program Logic Models
and Defining Metrics

Evaluation Planning: 3 Crucial Steps to Take *Before* Program Kick-off

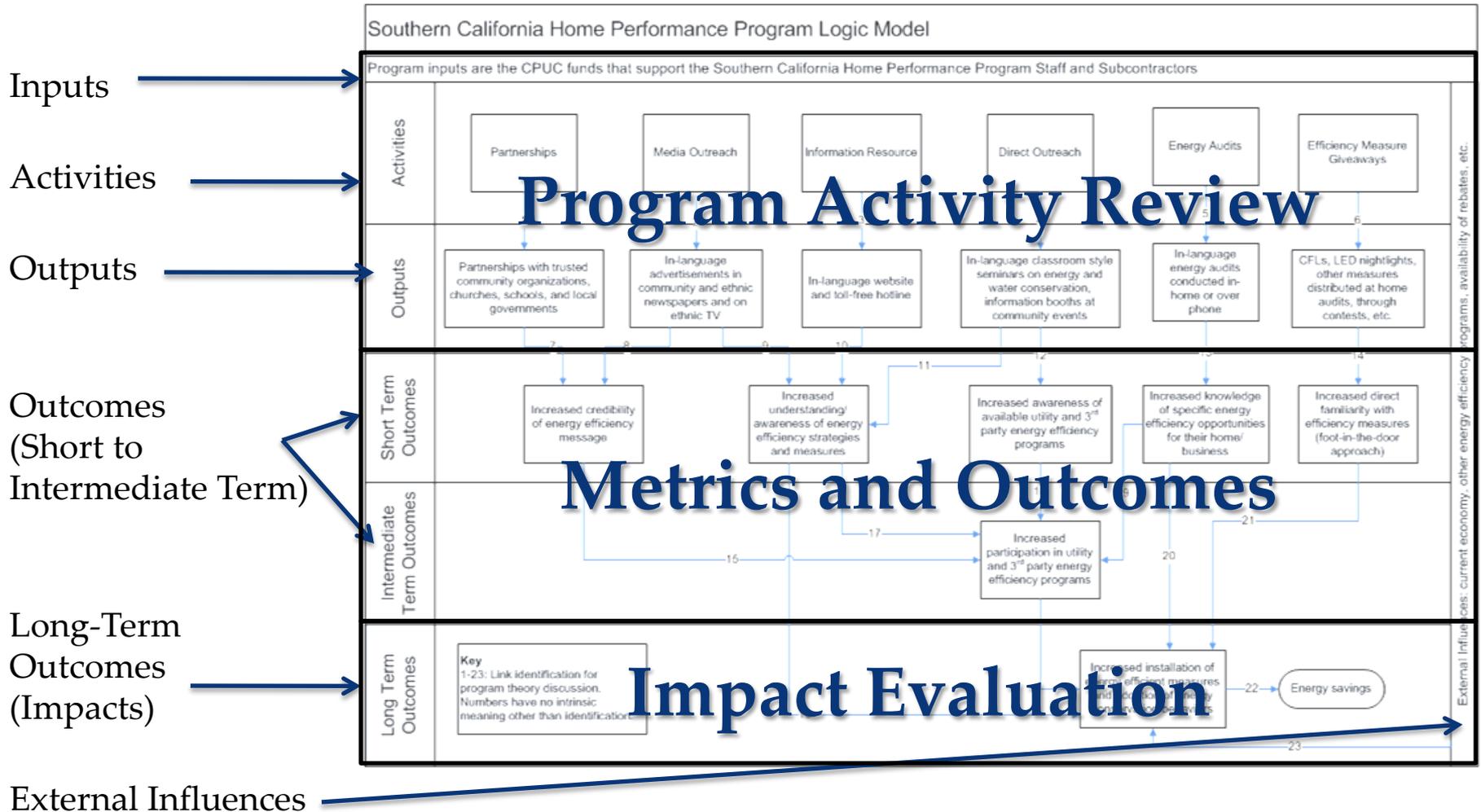


Step 1: Develop the Program Theory/Logic Model



External Influences →

Step 1: Program Theory/Logic Model's Links to Evaluation Strategies

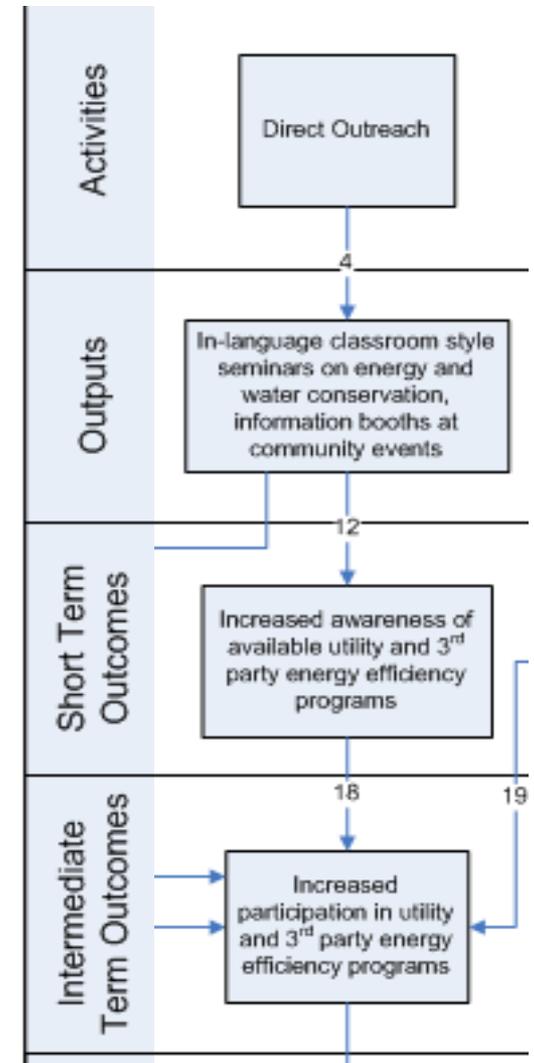


Step 2: Define Evaluation Metrics to Be Tracked

- Use the program theory/logic model to define specific metrics or progress indicators for each program activity
- Assess metrics at program outset (baseline) and throughout program implementation
- In the example to the right, metrics might include: number of seminars held; number of community event booths; awareness of EE programs, knowledge of EE opportunities; and participation in EE programs.
- Metrics can include outputs and outcomes.

Step 3: Conduct Baseline Research

- Measure the key evaluation metrics prior to program implementation, to track changes over time



Program Activity Review:

Measuring Program Outputs and Assessing Quality of Implementation

Measuring the program's direct outputs

- Outputs are actions taken directly by the program implementers
- Examples of a program's outputs:
 - Community outreach events held
 - Educational/motivational materials developed
 - Other communications (bill inserts, newsletters, emails, posters, etc.) developed
 - Financial incentives awarded
 - Etc.
- Outputs are typically measured by reviewing program records and expenses
- Evaluators should summarize the program's budget expenditures, timing of program activities, progress towards goals, etc., relative to the program design
- Goal is to provide a high level snapshot of program progress in time for course correction if necessary

Assessing effectiveness of program implementation

- **Benchmark program strategies** and messages against best practices in behavior change and social marketing; some good sources include:
 - Precourt Institute: <http://peec.stanford.edu/behavior/>
 - Community-based Social Marketing: www.cbsm.com
- Assess **actual program implementation** relative to program design, i.e., is the program being implemented as intended? If not, why?
- Interview program staff and review program implementation processes to identify **what's working** as well as **opportunities for program improvement**

Possible topics for review:

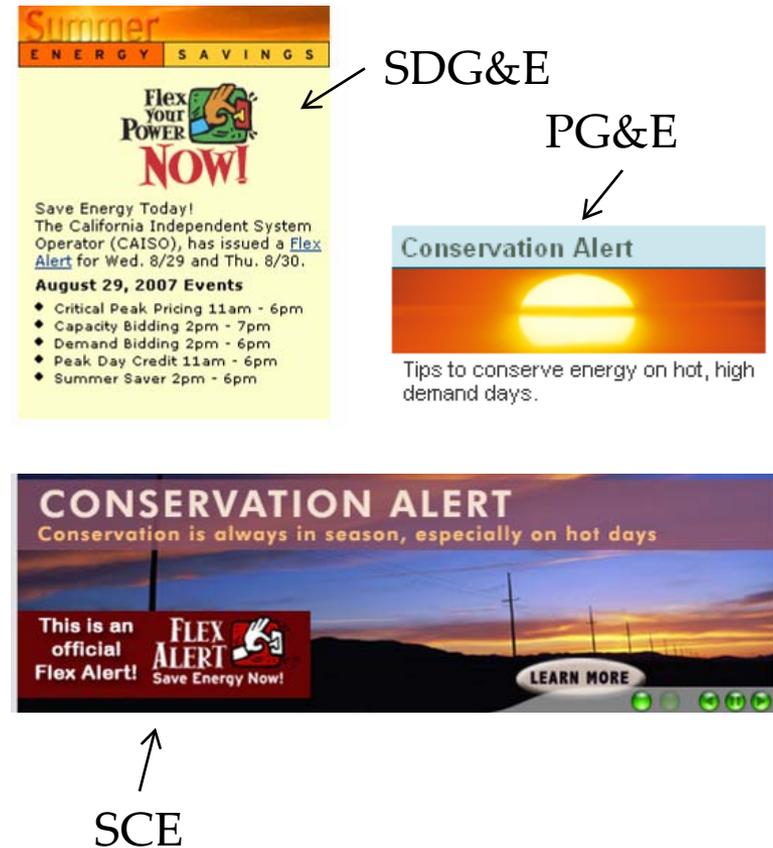
- Program management
- Staffing levels
- Staff skills
- Coordination between program staff and stakeholders
- Information tracking systems
- Reporting
- Quality control methods
- Target audience selection
- Message development/testing
- Outreach strategies
- Incentive delivery mechanisms
- Etc.

Does this program follow behavioral program best practices?

- Successful behavior change programs often:
 - Use **segmentation and barriers research** to inform program design
 - Design programs **with evaluation in mind**: baseline research, pilots, control groups
 - Select appropriate **marketing channels**
 - Empower **local change agents**
 - Invoke **social norms**
 - Meet consumers' desire for **quantification and prioritized recommendations**
 - Generate a sense of **enthusiasm and excitement**
- *Adapted from Navigant Consulting/Summit Blue's consumer behavior research meta-analysis conducted for the Northwest Energy Efficiency Taskforce. Full report available at <http://www.nwcouncil.org/energy/neet/workgroups/4/>.*

Flex Alert Evaluation Case Study: Program Activity Review

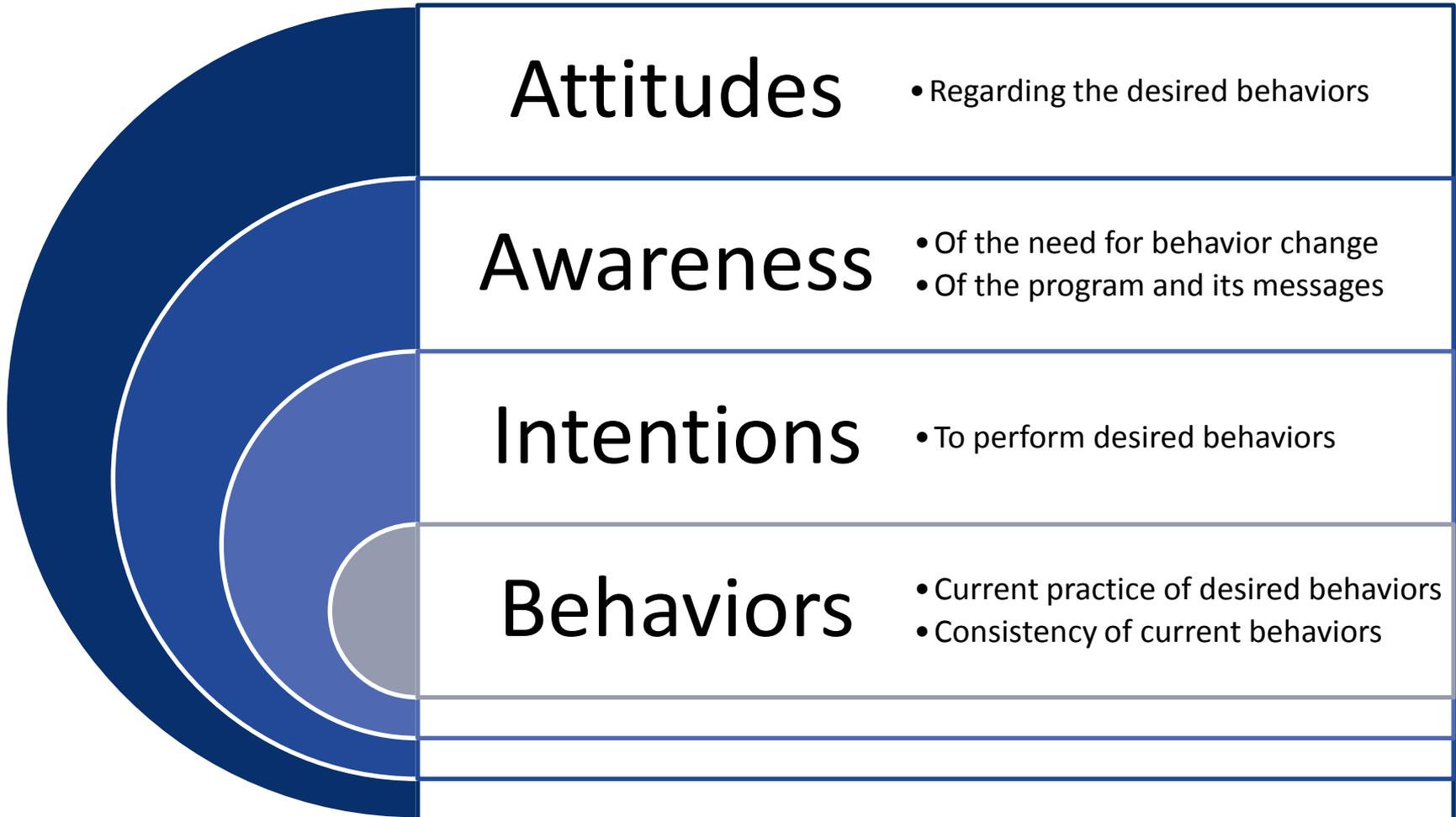
- Evaluation of California's Flex Alert campaign to promote electricity conservation on peak demand days
- Program activity review included:
 - Interviews with program implementers on processes and coordination
 - Review of media purchasing records
 - Focus groups to assess effectiveness of messaging and explore alternative target audiences
 - Review of press releases and website communications from multiple stakeholders for consistency and clarity of message
 - Providing recommendations on improvements to program design



Metrics and Outcomes:

Measuring Changes in Attitudes,
Awareness, Intentions, and Behaviors

Measuring the program's outcomes via surveys

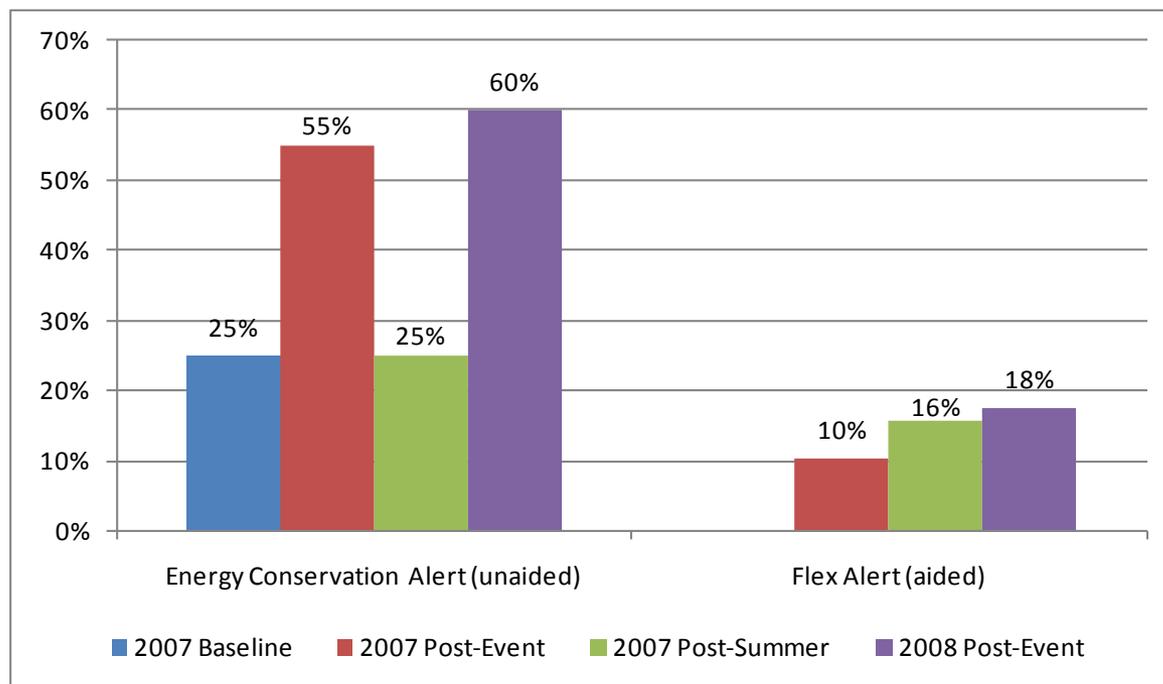


Measuring outcomes: surveys, media tracking, website analytics

- Test for unaided awareness of program and messages:
 - Start with open-ended questions about the program’s general message/promoted behaviors without mentioning the program by name
 - Provide a list of program names—including some dummy names—and ask respondents to select which names they are familiar with
- Collect demographic information to assess whether certain groups are more or less responsive to program messages, which can aid in future efforts.
- Use statistical tests such as t-tests to determine if data patterns are statistically significant; a software package such as SPSS makes this analysis easy
- Measuring outcomes may also involve analysis of traditional and/or social media “buzz”
- If program has a website, use a free tool such as Google Analytics to analyze volume and patterns of traffic to the website, particularly related to key events within the program

Flex Alert Evaluation Case Study: Metrics and Outcomes

- To assess outcomes of Flex Alert campaign, Californians were surveyed on recall of different Flex Alert messages and awareness of requested behaviors
- Surveys were conducted: 1) prior to the 2007 summer season (baseline survey), 2) immediately after a summer 2007 event, 3) at the end of the 2007 summer season, and 4) immediately after a summer 2008 event



Impact Evaluation:

Estimating Energy and Demand Impacts from Changed Behaviors

Measuring the program's impacts

Billing Analysis		Direct Metering	
Pros	Cons	Pros	Cons
<ul style="list-style-type: none"> •Accurate impact estimates for programs with many participants and a control group •Data is already being collected, just needs to be processed into useful format 	<ul style="list-style-type: none"> •Doesn't tell you <i>what</i> people did, only how much energy they saved •Need large sample sizes to be able to discern small impacts 	<ul style="list-style-type: none"> •Most accurate impact estimates •Provides some information even with small sample sizes 	<ul style="list-style-type: none"> •Expensive, especially if participants are unknown •Risk that the knowledge that they are metered may alter people's behaviors
Surveys + Engineering Analysis			
Pros		Cons	
<ul style="list-style-type: none"> •Provides richer information about <i>what</i> people actually did to save energy •Provides ability to assess possible other influences on behavior for attribution purposes 		<ul style="list-style-type: none"> •People can (intentionally or unintentionally) give inaccurate responses •Less statistically certainty of impacts, especially for programs that promote multiple behaviors 	
<ul style="list-style-type: none"> •Survey results could be calibrated with verification of a subset of respondents, to identify possible biases in self-reported actions/technical details of buildings/equipment. 			

Indirect impact analysis: surveys combined with engineering analysis

Identify Specific Actions Taken by Participant

- Word questions carefully to avoid framing the desired behavior as “the right answer” (i.e., social desirability bias)
- If program promotes multiple actions, consider prioritizing a few which are expected to account for the most significant energy/demand savings

Ask Contextual Questions to Assign Impacts to Actions

- Ask follow-up questions on building and equipment specifications; use questions appropriate to the level of the respondent’s technical expertise
- Ask follow-up questions about frequency and timing of actions (timing is especially important if peak demand savings are a goal of the program)

Attribute Savings to Program

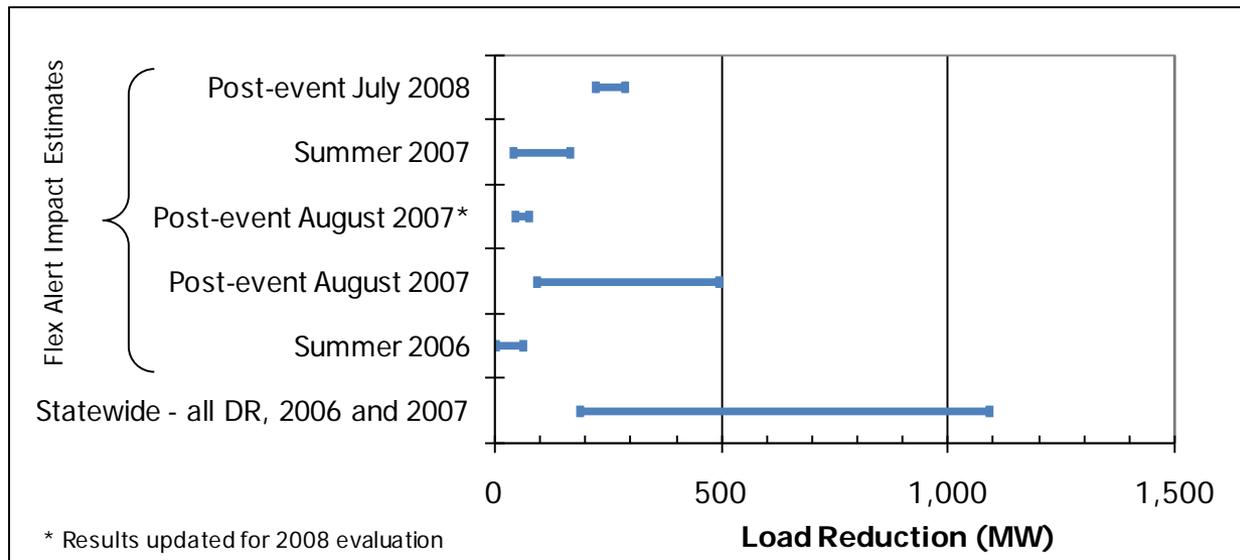
- Assess participants’ awareness/recall of the program’s messages *as well as other possible influences on their energy-related behaviors* to determine attribution of savings

Track Persistence of Program Impacts

- Track key metrics over time, including 1-2 years after program implementation if possible, to determine persistence of impacts.

Flex Alert Evaluation Case Study: Impact Evaluation

- Impact analysis focused on air conditioner and lighting behaviors promoted by the campaign
- Used assumptions about technical details that people are unlikely to provide accurately (e.g., bulb wattage); focused survey questions on specific actions taken and timing and frequency of actions
- Impact estimate of 2008 Flex Alert demand response event was 222 to 282 MW of peak demand reduction



Closing Thoughts

- Lack of evaluation research remains a barrier to behavioral programs
- Well-planned, systematically executed evaluations:
 - Provide feedback on program effectiveness and improve design/delivery
 - Aid in strategic decisions about future expenditures
 - Contribute to body of knowledge about behavioral programs
- Program theory and logic models are especially useful for behavioral programs in which there are often many steps between the program's direct outputs (which may be purely informational) and the expected impacts
- A theory-driven evaluation approach ensures that you are measuring the right metrics and evaluating the program in a systematic manner