Comparing Traditional Process Evaluation with the Developmental Evaluation Approach: Experiences with HVAC Programs as Part of a System for Change

Kara Crohn, Ellen Steiner, and Matt Galen, EMI Consulting

ABSTRACT

This paper explores potential uses of a developmental evaluation approach for supplementing traditional approaches to evaluating California HVAC programs. The authors consider developmental evaluation's unique attributes and discuss potential advantages and challenges associated with supplementing current practices with this approach. These discussions are grounded in examples drawn from recent evaluations. Traditional process evaluations ask questions such as: How are program goals reached and what improvements should be made? Traditional impact evaluations seek to document the value and persistence of program impacts attributable to program implementation. Developmental evaluation, in contrast, asks questions such as: To what principles should programs be held accountable, how do programs adapt to contexts in which they operate, and how do programs contribute to sustainable changes?

The authors submit that a developmental approach will improve evaluators' abilities to respond to technological changes and market complexities. In California, the HVAC industry and utilities are attempting to reduce energy use associated with HVAC equipment operation and maintenance through innovative energy efficiency programs and a statewide HVAC strategy. The industry must consider interwoven issues around equipment efficiency, quality installation, quality maintenance, and market demand. Yet current traditional program evaluations do not fully capture the complexity of the situation. These evaluations assess linear progress toward discrete program goals, rather than evaluate program efforts as part of a larger system working to fundamentally shift HVAC practices. Even if energy evaluators cannot fully adopt a developmental approach, aspects of developmental evaluation may provide significant value to the long-term impact of these programs.

Introduction

With HVAC energy consumption in the United States accounting for 30% of average summer peak-day loads and 22% of total electricity use (USEIA 2003, 2009), HVAC persists as a key focus for the California Long Term Energy Efficiency Strategic Plan (CPUC 2011). The industry knows it must work through inter-related issues around equipment efficiency, quality installation, quality maintenance, workforce education and training, and market demand. In California, utility-run energy efficiency programs dating back to the 1990's have tried to find ways to save energy through upstream energy efficient HVAC equipment stocking incentives, efficient equipment rebates, and contractor incentives for adhering to installation and maintenance best practices. More recently, the Western HVAC Performance Alliance (WHPA) has adopted a systems perspective that engages key stakeholder groups in an attempt to raise the bar for maximizing reductions in energy costs through the selection, installation, and maintenance of HVAC systems.

Traditional program evaluation approaches are not suited for tracking progress toward dynamic systems-level outcomes. Until recently, program evaluations supporting this strategy often assessed linear progress towards singular program goals, rather than considering program

activities and outcomes as part of a larger system working to shift fundamental HVAC industry practices. Current program evaluations have embraced more of a systems perspective, yet there is room to improve. To better capture these complex, systems-level shifts, we should consider designing evaluations of innovative HVAC programs through the developmental evaluation approach (Patton, 2010). Developmental evaluation aims to aid continuous improvement and rapid adaptation to a changing market. A traditional process evaluation of a relatively static program asks questions like: What are the program goals, how are they reached, and what improvements should the program make? Developmental evaluation is suited for innovating in dynamic environments with multiple, complex, and evolving goals, which is often the case in new HVAC programs. Examples of developmental evaluation questions might include: To what principles should a program adhere and be held accountable, how do the HVAC programs adapt to the context in which they operate, and how do the HVAC programs contribute to a sustainable change in the HVAC market?

This paper compares traditional and developmental evaluation approaches, using California Investor-Owned Utility (IOU) HVAC program evaluations as illustrative examples. The goal of this exercise is to prompt discussion and debate about how the energy program evaluation industry could successfully implement this vitally needed approach. In this paper, we first discuss this need for evaluation designs that support adaptive management. Much of this need is rooted in the quality of uncertainty posed by market transformation initiatives as compared with more traditional resource acquisition programs. We then provide an in-depth discussion of the key differences between the developmental evaluation approach and more traditional formative and summative evaluation approaches. Real-world examples drawn from traditional evaluations of California HVAC programs will serve as a starting point for illustrating how these evaluations could be different using a developmental evaluation approach. Finally, based upon this comparative assessment we summarize the challenges and opportunities in using a developmental evaluation approach.

The Need for Evaluation Designs That Support Adaptive Management

In recent years, energy efficiency evaluation efforts have focused at the program-level on three main types of evaluation activities: (1) market characterizations and assessments that help describe market conditions, adoption rates of the technology, or behavior of interest, (2) process evaluations describing and identifying opportunities to improve program theories, implementation processes, and marketing approaches, and (3) impact evaluations seeking to quantify energy savings and market effects for individual programs. In addition to determining whether or not a program works, program managers and regulators are increasingly seeing the value of understanding how and why programs work. In recent years, for example, energy efficiency programs are more often required to document program theories through logic models. This encourages utilities and other energy efficiency program providers to bring evaluators in at the program planning stages instead of waiting until the program is nearly done. Program staff can work with evaluators to carefully consider how they can best design program activities to reach desirable outcomes, how those outcomes would lead to desired short- and long-term goals, and what external forces would affect their ability to accomplish these outcomes and goals. Involving evaluators early in the program development process also offers an opportunity to design the evaluation to test the assumptions and connections presented in the program logic model, and ensures the program's capacity to collect necessary evaluation data throughout the program design and implementation processes. This prepares programs to receive feedback about

the program throughout its design and deployment stages so that program staff can make any needed changes before seemingly small issues bubble up into fatal flaws. In this way, our evaluations are becoming more responsive to shifting program needs and are contributing to ongoing program design and redesign.

However, this does not go far enough. These evaluation approaches often assume a simple linear route to market development and a linear connection between program activities and outcomes. Logic models are often treated as static documents; as the stake in the ground against which the programs' ability to affect desired outcomes are assessed. Yet many market transformation-oriented energy efficiency programs would be doing themselves a disservice to plunge a stake in the ground too deeply. These programs try to induce change in a much broader dynamic system, with multiple forces influencing if, when, and how changes take place. Markets can shift before full market transformation takes place, rendering a program's theory of change out of step with current conditions (NMR 2013). Programs must either update the activities and goals on those logic models on a regular basis to reflect current sources of influence, or they must represent the relationships between the program's activities and desired market conditions more dynamically with alternative market models.

Recent research by NMR advised that evaluations of market transformation programs require a different approach to evaluation because, unlike the more hands-off approach of traditional evaluations, "it is essential for the success of MT programs that program planners, designers, and implementers work hand-in-hand with evaluators," to facilitate program development (NMR 2013 p. 24). According to the study's suggestions, logic model development can be improved by first determining the transformation theory and mechanism by which the program intends to influence a market. This market transformation narrative and picture should be placed in a particular time frame that includes an exit or (perhaps more likely) a transition strategy with relevant indicators such as, "market actors are in the position to continue to facilitate adoption of the product or service," "reverting to earlier equipment would be costly," and "more efficient codes and standards have been adopted" (NMR 2013 p. 19). The study also recommends continuing to monitor key market indicators after the market transformation has occurred to determine if the transformation persists. This speaks to the need for market transformation programs to be designed in a manner that allows for rigorous accountability for program activities and outcomes while also allowing for enough flexibility to respond to market demands.

If a market transformation program is designed in a way that facilitates adaptation, as the research by NMR suggests, then the program will likely need to use a management approach that allows for engagement in continuous improvement cycles, and also anticipates and responds to market changes. Such an approach may fundamentally alter the design of the program. An evaluation of such a program (or portfolio of programs) must also be able to capture these changes and assess progress towards the end goal. The developmental evaluation approach provides guidance on creating this type of responsive, rigorous evaluation.

Overview of Developmental Evaluation

Developmental evaluation supports an adaptive management approach. In contrast, traditional energy efficiency evaluation generally focuses on ex-post impacts, with supporting roles for process and market evaluations. Michael Quinn Patton coined the term "developmental evaluation" to describe an evaluation approach that could support the development of innovations occurring in dynamic, complex environments. In these environments, knowing,

"what to do to solve problems is uncertain and key stakeholders are in conflict about how to proceed," (Patton 2011 p. 1). Examples of innovations appropriate for this developmental approach include, "new projects, programs, products, organizational changes, policy reforms, and system interventions," (p.1). Patton further describes the five purposes and uses of developmental evaluation that support more of an adaptive management approach to developing programs. These include: (1) "ongoing development," (2) "adapting effective general principles to a new context," (3) "developing a rapid response in the face of a sudden major change," (4) "preformative development of a potentially scalable intervention," and (5) "major systems change and cross-scale developmental evaluation, providing feedback about how major systems change is unfolding, evidence of emergent tipping points and/or how an innovation is or may need to be changed and adapted as it is taken to scale" (Patton 2011 p. 21).

In terms of HVAC energy conservation and efficiency programs, uses (1), (2), and (4) are most likely applicable. For the first purpose of ongoing development, Patton describes using developmental evaluation to help a program or strategy adapt to new conditions in the system. In this sense, the purpose of some evaluations of California IOU HVAC programs or statewide strategies could be to document changing market conditions and determine how best to adapt the strategy to meet market demand while working towards energy efficiency goals (EMI Consulting, forthcoming and 2012a). For the second purpose, Patton explains that developmental evaluation can be used to help determine how to adapt general principles that were effective in one situation to a new situation. We can imagine conducting an evaluation tasked with learning what underlying principles guided a successful HVAC program and working with program managers of another HVAC program to develop a similar pilot HVAC program. Metrics could test the extent to which programs adhere to design principles, as opposed to – or in addition to – tracking specific activities executed, or specific amounts of energy saved. For the fourth purpose, Patton describes how developmental evaluation can be used to explore the possibility of transforming a vision or idea into a program that is operational and potentially scalable. In HVAC energy efficiency program evaluation, there is an easy corollary to helping pilot programs evolve to the point where they can be scaled up to stable programs, where they are then evaluable using more traditional approaches.

Developmental Evaluation vs. Traditional Evaluation: California HVAC Quality Installation and Maintenance as an Example

Adapting Patton's comparison of traditional and developmental evaluation (see Table 1), we can identify opportunities and potential limitations to adding a developmental approach to evaluating California IOU HVAC programs. The program menu in California is moving beyond retrofit programs to target levels of deeper energy efficiency that may be gained through quality installation and maintenance of HVAC equipment. This is a *very* difficult market to impact given the variety of stakeholders involved, low customer demand for high quality services, variable knowledge and skill levels of the workforce (Don Vial Center 2011; EMI Consulting 2012b), competition driven by price, and a lack of code enforcement. In sum, the current situation for California HVAC is driven by a variety of industry stakeholders who must collaborate over a long period of time to transform multiple, interrelated aspects of the market in order to realize energy savings and ensure those savings persist.

California IOU HVAC quality installation and quality maintenance programs aim primarily to help contractors shift their business models. The utilities provide incentives to offset the additional cost of installing and maintaining a system to meet ACCA/ASHRAE standards.

Contractors have to agree to follow specific installation and maintenance steps, and the utilities perform desk and onsite verifications to confirm that the technicians are performing the work correctly. The utilities initially provided training to contractors and technicians to show them what tasks needed to be performed to participate in the program and how to record information about their installations or maintenance tasks. However, the program staff quickly found that many technicians and contractors did not know fundamental job skills. While this "how to" type of training would ideally come through HVAC training programs, the programs decided to help fill a need for improved technician skills by mentoring and training technicians involved in the program. This type of change goes beyond making program improvements and into the realm of redesigning key aspects of the program.

Table 1 provides an overview of our comparative assessment of the evaluation approaches for these programs, considering the (1) purpose and situation of the evaluation, (2) focus and target of the evaluation, (3) modeling, methods, and approaches to complexity, (4) roles and relationships, and (5) evaluation results and impacts. This is necessarily a simplified typology, but serves to illustrate critical differences in approach and intended outcomes from the evaluation activities, as discussed in more detail below.

Purpose of the Evaluation

California IOU HVAC program evaluations generally lean towards the traditional view of evaluation's purpose – to use a combination of formative and summative evaluation approaches that seek to help HVAC programs make incremental improvements, draw conclusions about whether the program was implemented with fidelity to the pre-planned design, and verify if the program produced sufficient energy savings in a cost-effective, scalable manner.

A developmental approach is better suited for providing real-time feedback about a program's contribution towards building lasting changes to HVAC contractors' business practices. In the first few iterations of the programs, the staff need time to determine which combinations of support mechanisms are sufficient to spark change, determine whether that support should come from the utility or in collaboration with another market actor(s), and make sure the contractors can thrive under the new model. As the program evolves, the evaluations should adapt to provide quick feedback not only on new efforts, but also look at the effects on other industry partners that play key roles in advancing standards of practice. Evaluations should also raise awareness about the importance of quality installation and maintenance.

Focus of Evaluation

Developmental evaluation focuses on documenting systems change rather than on the unique contributions of isolated programs or program components. Outcomes are emergent, not pre-determined. The evaluation seeks to provide timely feedback for ongoing program development, much like the role of evaluation in an adaptive management practice. A working assumption underlying traditional evaluation approaches is that the program is trying to reach what Patton refers to as an *engineering* sense of resilience - one that "focuses on efficiency, control, constancy, and predictability under conditions of low uncertainty." This can be contrasted with an *ecosystem* sense of resilience that "focuses on persistence, adaptiveness, variability, and unpredictability under conditions of high uncertainty" (Patton 2011 p. 199). Developmental evaluation is more aligned with the idea of achieving ecosystem resilience, in that it is not concerned with determining the precision with which program outcomes can be predicted, but rather the degree to which programs are able to adapt to meet emerging needs.

Table 1. Developmental evaluation vs. traditional evaluation: California HVAC Quality Installation (QI) and Quality Maintenance (QM) as an example

Comparative Attribute	Traditional (Current) Approach	Developmental Evaluation Approach
Purpose of Evaluation	Inform incremental changes, assess compliance with and scalability of design, and quantify market transformation and energy savings impacts from stable QI and QM programs with intervention plan that is concrete, measurable, and unlikely to change within a few years.	Provide quick, real-time feedback to QI and QM program staff and key industry stakeholders (e.g., WHPA, HVAC training providers) to inform the iterative adaptation of program's design and theory of change as it relates to shifts in QI and QM customer demand, contractors' ability to sell QI and QM, and availability of contractor training.
Focus of Evaluation	Measure (1) effects attributable to programs on adoption of QI and QM practices and certainty of energy savings, (2) validity of engineering estimates, and (3) operational efficiency.	Document and help program staff and stakeholders reflect on QI and QM programs' roles in shifting HVAC industry towards business practices that will lead to long-term savings.
Modeling, Methods, Approach to Complexity	Evaluation designed to hypothesize and test links between QI and QM logic model activities and outputs, and outputs and outcomes using counterfactual to test for attribution of program effects on broader HVAC industry problems.	Evaluation designed from systems perspective to track emergent decisions, anticipated and unanticipated program outcomes, and interdependencies using tools such as dynamic theories of change, market transformation models, and social network analysis.
Roles and Relationships	Objective 3rd party evaluator works minimally with program staff and program participants to gather data about the program and is accountable to the CPUC, IOU internal evaluator, and QI and QM Program Managers to provide an independent, written assessment of the QI and QM programs.	Objective 3rd party evaluator is accountable to CPUC, IOU internal evaluator, QI and QM Program Managers, and industry stakeholders. He/she also works closely with program implementers and stakeholders to ensure long-term success of integrating evaluation findings into a continuous cycle of program management improvement.
Evaluation Results and Impacts	Formal evaluation report focuses on replicability of the QI and QM programs, describes validity of data collection and analysis, and details findings leading to recommendations about program continuation.	Rapid, accessible, interactive reporting focuses on supporting learning, communication, and knowledge transfer among stakeholders to facilitate evolution of QI and QM programs in service of HVAC goals.

We can imagine an HVAC program seeking engineering resilience as one that focuses on providing incentives for individual measures that have demonstrated the highest, most reliable energy savings while controlling for other variables. Such programs tend to estimate program impact by adding up the individual contributions. In comparison, a program seeking ecosystem resilience would focus on funding a mix of measures, marketing activities, technician training, and reward systems that in combination would likely have the best chance of shifting market trends in favor of buying, installing, and maintaining more energy efficient HVAC equipment in the near future. If a lack of training was identified as a key barrier, an HVAC program that demonstrated ecosystem resilience would be able to modify its program design to include mentoring technicians. Such programs tend to estimate program impact by assessing influence on overarching shifts in practice.

As the industry seeks to better understand and improve the way in which HVAC installation and maintenance work is conducted (as opposed to focusing on improving the efficiency of a piece of equipment), the level of complexity and interdependency increases, and the certainty, predictability, and persistence of the desired changes decreases. Some market transformation programs approach program design with more of an adaptive management approach in mind, such as Southern California Edison's HVAC Optimization program, which uses Six Sigma tools as a foundation to drive continuous improvement for responding and adapting to its market.

Evaluation questions answered through a developmental evaluation approach are more in line with the types of questions market transformation programs need to answer. HVAC energy efficiency programs might include questions we normally associate with market assessment and with pilot program development, such as (see Exhibit 2.2. in Patton 2011 p. 44-47): What is the baseline understanding of where the HVAC industry stands in terms of reducing HVAC system energy consumption, or where are the opportunities for key industry stakeholders to collaboratively tackle known issues like inconsistent installation and maintenance practices? Thus far, mostly discrete, separately-funded studies attempted to address aspects of these example questions.

Three such studies were conducted in 2012. The first described the extent to which technicians followed residential quality maintenance guidelines (EMI Consulting 2012a). The second study, a needs assessment, characterized the current state of 20 important HVAC roles in the industry according to the tasks, knowledge, skills, and abilities required for the jobs (EMI Consulting 2012b). The third study discussed important industry values and key stakeholders as it described the WHPA's progress to date working with HVAC market actors to accomplish California Strategic Plan HVAC goals (EMI Consulting 2012c). Taken together the studies shed light on a set of interwoven issues the HVAC industry must work through to bolster the knowledge, skills, and abilities of HVAC industry actors such that they can implement new industry standards.

As a set of HVAC energy efficiency programs begins developing, the evaluation can support development of new models that recognize the interconnections among desired changes in HVAC equipment stocking, installation, and maintenance practices. The evaluation of pilot programs can help create testable boundaries and identify key issues that need to be addressed as the best model is selected to formalize HVAC program plans, such as: What do the results of rapid feedback reveal about progress towards changing HVAC equipment stocking, installation, and maintenance practices? What criteria emerge as important to determine what is "working" and "not working"? How can the evaluation adapt to track those criteria across program efforts and industry collaborations, like WHPA? A rapid-feedback process evaluation of the HVAC

Optimization program took a snapshot of program changes underway and identified potential gaps in program processes that staff could address as they instituted the next round of program changes. It was challenging to document the "current program" in logic model and process map format as the program was shifting in response to feedback from contractors on previous program changes during the evaluation. However, the evaluation process – which reflected aspects of a developmental approach – offered program staff an opportunity to revise the logic model and reflect on some key assumptions undergirding the program as they planned their next wave of program changes.

Modeling, Methods, and Approaches to Complexity

Developmental evaluation designs are more likely to use complex system maps rather than linear logic models. These maps track emergent interconnections and fluctuations in the strength of those ties rather than assume stable and linear relationships. Without relying on the assumption that the program will remain stable over time, the evaluation could respond to unfolding events and could detect causal patterns in the data through retrospective analysis. Performance metrics are likely to be identified and tracked as they emerge, and may change to allow evaluators to gather data that will inform the next innovation decision point. Documenting these decision points and plausible alternatives is important for understanding implications of decisions and sources of influence on decisions. As such, documenting unexpected and unanticipated events is important. The usefulness of the evaluation is likely to depend upon its contextual relevance and the ability of decision makers to take action based on the process or findings.

Currently, California IOU HVAC programs use linear logic models to describe how their programs work. These logic models do a decent job of explaining the activities the program performs and the desired short, intermediate, and long-term outcomes. However, what the logic models often neglect to do is visualize how the programs work with other stakeholder groups and explain how they work together towards shared goals. Stakeholders are often relegated to a small box in a corner of the logic model titled, "Other Market Actors." A traditional evaluation perspective might try to hold constant the effect of the other market actors in order to attribute a portion of the program effect to the program activities, such as utility-driven training efforts. A developmental evaluation would seek to demonstrate how relationships with market actors contribute (or do not contribute) to developing a new approach to overcoming a market barrier. For example, for a combined effort to roll-out hands-on Standard 180-aligned training across community college, distributor, certification, and utility-run classes, the evaluation would consider how the utility collaborates in its achievement (or lack of achievement) of the training goals.

A key challenge to any market transformation energy efficiency program is determining how to track energy savings while remaining nimble enough to make substantial program changes in response to market shifts (or anticipated shifts). This challenge is not new, but resolving this challenge requires finding a middle ground between pre-determined and emergent evaluation design.

Let's take the SCE Quality Installation Program as an example of how an evaluation might strike such a balance in practice. Fundamentally, the program wants to see an increase in the number of units technicians install according to ACCA/ASHRAE standards and convert contractors' business models to follow this approach without program incentives and training. It seems plausible that a traditional longitudinal pre-post verification study could be built into the

HVAC program design with appropriate sampling requirements without compromising the program's ability to add to or radically change aspects of their program, like incentive structures, training, and mentoring. As part of the developmental evaluation approach, changes to the program would need to be tracked and taken into account during analysis of savings. Short-term studies could occur in parallel to the larger longitudinal study to provide rapid feedback on pilots of new activities. A smaller study might, for example, study new training initiatives that require exploratory qualitative methods to effectively document decisions, emergent benefits and unanticipated outcomes. If attribution remains a goal for determining whether an initiative is effective, using an evaluation approach such as the Kirkpatrick/Phillips model for assessing training program return on investment presents layers of outcomes that can be tracked (Bailey). With this type of evaluation model, the information can be fed back to decision makers about each outcome layer in time for them to adapt their approach quickly in response to findings.

Roles and Relationships

The role of the evaluator in a developmental evaluation is less judge and jury of a singular program and more akin to an observer, questioner, or facilitator who is "part of a development team whose members collaborate to conceptualize, design, and test new approaches in a long-term, ongoing process of continuous development, adaptation, and experimentation, keenly sensitive to unintended results and side effects. The evaluator's primary function in the team is to infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection and decision making in the developmental process" (Patton 2011 p. 1-2). The emphasis is on evaluating a program or organization's efforts to deal with a societal or systemic issue as opposed to focusing on a single program's efforts to address one aspect of the bigger issue, largely in isolation from other discrete efforts.

Energy efficiency policy makers tend to value arms-length approaches to evaluation of singular programs held to program-specific goals. Often in California, different evaluation firms must conduct process and impact evaluations out of fear that evaluators will be coopted to support the program staff or implementers' agendas and bias the findings if they play a more active role on the program team. Yet, we evaluate market transformation programs that are trying to respond to – and speed up – changes in dynamic, complex markets. These programs need accurate feedback, and they need it quickly and frequently in order to respond effectively. In our efforts to ensure attribution, objectivity, and precision, we risk stifling the very type of innovation we hope to support by encouraging programs to act more independently than they should.

Evaluation Results and Impacts

Evaluation results of traditional evaluations tend to be reported formally, in a scholarly voice intended to maximize credibility. The evaluation process and reporting style tend to focus on ensuring credibility of results by adhering to methodological standards. The developmental evaluation approach adheres to the high methodological standards of traditional evaluation, but emphasizes reporting on effective principles and requirements for success, potentially testing adherence to those principles or offering guidance for applying principles in other contexts. The results should provide real-time feedback to the program in an active, present voice that engenders a desire to learn and adapt. In developmental evaluation, stakeholder capacity is built throughout at all stages of the evaluation to facilitate effective reporting. Evaluators work with

stakeholders to build evaluative thinking into the critical reflection phase of an adaptive management process (plan, implement, gather data, reflect … repeat).

Working with California IOU HVAC programs, EMI Consulting has provided both traditional written evaluation reports and rapid-feedback products, such as interactive presentations and models that visualize results and foster conversation. For accountability purposes, the programs tend to require written reports that are vetted by CPUC and interveners. However, the program staff and stakeholders closest to the program decision-making process benefit more from presentation and discussion of findings, well before a long written report can be produced. These rapid feedback mechanisms can be designed to surface opportunities or threats that may warrant a major program redesign rather than to regularly "check-in" on an established program.

Summary: Opportunities and Challenges to Implementing Developmental Evaluation

Using the California IOU HVAC market transformation programs as an example, we can see some important opportunities and challenges to implementing a developmental evaluation approach. The HVAC industry is evolving and utility programs are trying to fulfill two roles: (1) generating reliable energy savings while (2) pushing the industry to transform the way in which HVAC work is performed. These goals are complementary in that an increase in consistent high-quality installation and maintenance practices should lead to increased energy savings over time. The state supports market transformation programs that will inherently adapt to changing market conditions to speed up the adoption of good work practices as well as more efficient equipment. Traditional evaluation approaches are not typically nimble enough to track the effects of such evolving program designs that operate in a broader dynamic market. Developmental evaluation offers alternative ways to understand and facilitate market transformation efforts.

As with most evaluation theories, some aspects of the approach are more readily applied than others. With California IOU HVAC programs, there are many opportunities to apply developmental evaluation practices, including designing rapid-feedback studies, developing models that characterize the program's role in broader market dynamics, and asking evaluation questions that emphasize replicability of key principles necessary to transform the market. Even presentation of information can be done with program evolution in mind versus summative judgment of worth or merit. Further, the scope and purpose of market characterization studies and process evaluations could be altered to evaluate efforts across a portfolio of programs.

There are hurdles to implementing a developmental evaluation approach with energy efficiency programs. Two of those challenges center around the role of the consultant and the need to attribute energy savings to utility efforts. In developmental evaluation, the role of the evaluator is that of a consultant who builds a long-term relationship with the client. This relationship is important in order for the evaluator to be a trusted member of the design team, present with the purpose of prompting other members to reflect on the current innovations underway and use evaluation findings to make decisions about the program's path forward (Rey, Tremblay, and Brousselle 2013). In California, internal IOU evaluators are allowed to interact freely with program staff. However, a close, long-term relationship between external evaluators and program staff would be frowned upon and considered suspect because the external evaluator would not be maintaining enough distance to be "objective." In order to overcome this perceived conflict of interest, California regulators would need to reconsider how they assure accountability and objectivity in external evaluation. For example, regulators could allocate

some evaluation resources towards periodic meta-evaluation, a process akin to peer-review by a third-party evaluator, to verify that the developmental evaluation approach was (1) adhering to the Guiding Principles (AEA 2004) and Program Evaluation Standards (Yarborough et. al. 2011), and (2) resulting in an improved program design. If developmental evaluators do their job well, they will provide accurate, timely feedback and facilitate discussion of the results among the design team. Evaluators themselves, though, should not be the program designers.

The second hurdle, the need to attribute energy savings to utility efforts, poses a challenge to the way we conceptualize attribution. From a systems perspective, attribution of effect to one component (e.g., a utility program) is meaningless because those components are not static and are intertwined with other components (e.g., broader industry initiatives, market demand, technician training); one cannot function fully without the others. However, utilities need to be able to count on energy savings generated from energy efficiency programs in order to forecast load estimates accurately. Utility programs are usually required to contribute reliable energy savings estimates. Estimating those savings likely requires tracking savings estimates in a consistent way over extended periods of time, which is more consistent with traditional impact evaluation approaches. In contrast, a developmental approach may be better positioned to identify and estimate spillover, other market effects and persistence of energy savings. In our HVAC example, we can see how this tension plays out. The HVAC quality installation and quality maintenance programs are under scrutiny to demonstrate energy savings, yet their successes are heavily tied to the successes of efforts outside of their direct control, including the industry's efforts to (1) reach agreement on how to perform the tasks outlined in the standards to maximize energy savings, (2) train technicians to install and maintain systems to maximize energy savings, (3) characterize (and then drive) market demand for higher-quality HVAC installation and maintenance, and (4) ensure HVAC systems are right-sized and installed properly.

To overcome this hurdle, regulators would likely need to reconsider how and when they account for energy savings. For example: Would it be possible to track energy savings at a portfolio-level that aligns with the state's HVAC strategy? Or, if the program theory is unlikely to stabilize for more than a few years, could we hold individual programs accountable for non-energy impacts, such as short- and intermediate-term market transformation indicators, at least for a period of time long enough to allow program theories to stabilize? This may entail tracking actions of other industry players that would indicate steps towards the desired change, as defined by market transformation indicators. Perhaps evaluators could help program staff identify program-level measures of successful adaptation while monitoring overall portfolio-level energy savings?

Developmental evaluation is not right for all situations, yet it works well as part of an adaptive management process and should be used to support market transformation programs' efforts to make meaningful innovations while tracking progress towards long-term goals.

References

AEA (American Evaluation Association). 2004."American Evaluation Association Guiding Principles for Evaluators," http://www.eval.org/p/cm/ld/fid=51

Bailey, A. No date. "The Kirkpatrick/Phillips Model for Evaluating Human Resource Development and Training," Learning Designs Online. Accessed on May 14, 2014: http://www.busgurus.ca/media/pdf/the-kirkpatrick-phillips-evaluation-model-en.pdf

CPUC (California Public Utilities Commission). 2011. *California Energy Efficiency Strategic Plan January 2011 Update*. http://www.cpuc.ca.gov/NR/rdonlyres/A54B59C2-D571-440D-9477-3363726F573A/0/CAEnergyEfficiencyStrategicPlan_Jan2011.pdf

(Don Vial Center) Don Vial Center on Employment in the Green Economy. (2011). *California Workforce Education and Training Needs Assessment for Energy Efficiency, Distributed Generation, and Demand Response.* http://www.irle.berkeley.edu/vial/publications/WET_Part1.pdf

EMI Consulting. (forthcoming). *HVAC Commercial Customer Decision-Making Study.* Prepared for: Southern California Edison Company.

EMI Consulting (Energy Market Innovations, Inc.). 2012a. *California HVAC Contractor & Technician Behavior Study.* Prepared for: Southern California Edison Company.

EMI Consulting (Energy Market Innovations, Inc.). 2012b. *HVAC Educational Needs Assessment.* Prepared for: Southern California Edison Company.

EMI Consulting (Energy Market Innovations, Inc.). 2012c. *WHPA Progress Report*. Prepared for: Southern California Edison Company.

NMR (NMR Group, Inc.). 2013. *A Review of Effective Practices for the Planning, Design, Implementation, and Evaluation of Market Transformation Efforts.* Prepared for: Pacific Gas & Electric, San Diego Gas & Electric, Southern California Edison, and Southern California Gas. http://www.calmac.org/publications/FINAL_NMR_MT_Practices_Report_20131125.pdf

Patton, M.Q. "Michael Quinn Patton on Developmental Evaluation." AEA365 (email blog), *American Evaluation Association,* received Sunday April 27, 2014. http://aea365.org/blog/de-week-michael-quinn-patton-on-developmental-evaluation/

Patton, M.Q. 2011. *Developmental Evaluation.* New York: The Guilford Press.

Rey, L., Tremblay, M., and Brousselle, A. 2013. "Managing Tensions Between Evaluation and Research," *American Journal of Evaluation* 35 (1).

Yarborough, D., Shula, L., Hopson, R., Caruthers, F. (2011). *The Program Evaluation Standards, 3rd Edition.* Thousand Oaks, CA: Sage Publications.