

Evaluating the Evaluation Method: A Retrospective on Embedded Evaluation

Shannon Kahl, ILLUME, an E Source Company, Boulder, CO

Liz Kelley, ILLUME, an E Source Company, Boulder, CO

Pace Goodman, ILLUME, an E Source Company, Boulder, CO

ABSTRACT

Evaluation methods meant to provide early and ongoing feedback, often referred to as embedded evaluation, can provide great value to a program, giving program staff the opportunity to adjust program delivery in real time. But is embedded evaluation a cure-all for evaluation pain-points, or does it only work well when applied in the right circumstances or when done with best practices? The truth is that there are shades of gray and there are circumstances and approaches to embedded evaluation that help ensure programs realize the benefits of embedded evaluation, like having access to rapid and accurate evaluation results, while mitigating the risks.

Successful embedded evaluation takes careful planning and a shift away from some of the practices of traditional, retrospective, or comprehensive, program evaluation. This paper draws from six case studies of embedded evaluations including two pilots in the west, a statewide program in the Northeast, a pilot in the Northwest, and two new offerings introduced by a utility in the Southeast to explore strategies to respond to or avoid some of the common challenges with embedded evaluation. These include a lack of awareness about what to expect with an embedded evaluation among utility and implementation contractor (IC) staff, misalignment on how and when recommendations will be provided and acted on, when to anticipate and how to right size evaluations with lower-than-expected enrollment or other changing program needs, and challenges receiving program and/or usage data required to complete evaluation tasks.

Our paper provides strategies to navigate the challenges that can come with trying to provide meaningful results through embedded evaluation.

Introduction

The term “embedded evaluation” has been used throughout our industry (and others) and may have different meanings to different practitioners. For the purposes of this paper, our definition of embedded evaluation aligns with the 2025 – 2027 Massachusetts Statewide Energy Efficiency Draft Strategic Evaluation Plan (Massachusetts Program Administrators 2024), which describes *embedded* evaluation as a process where evaluation activities are integrated into ongoing program implementation to provide a continuous feedback loop to the program team. While evaluation activities conducted as part of an embedded evaluation are often the same as those conducted in a more traditional retrospective, or comprehensive, evaluation, the purposes and priorities may differ. For example, findings and recommendations from an embedded evaluation tend to be more formative in nature and evaluators may prioritize providing rapid feedback over achieving specific confidence and precision. In contrast, the Massachusetts Statewide Energy Efficiency Draft Strategic Evaluation Plan describes *comprehensive* evaluation as the traditional evaluation style that often requires at least 9 months to design and execute. After all tasks are complete, the evaluation team then produces a final report aimed at assessing the program’s overall effectiveness and impact (Massachusetts Program Administrators 2024).

Evaluators can leverage embedded evaluation methods to provide ongoing or even real-time feedback as part of a process and/or impact evaluation, including measurement and verification (M&V).

However, successful embedded evaluation requires a paradigm shift, changing how we think about evaluation from how we scope evaluation tasks and work with implementation contractors, to procurement, to relationship building across teams, to the timing and structure of methods, and sharing results. It requires a much more adaptable infrastructure on both the client and evaluator side than more traditional retrospective program evaluation.

Drawing from our evaluation team’s experiences with these embedded evaluations and others, this paper explores strategies for successful embedded evaluation including how it can complement a more traditional evaluation and how to respond to or avoid some of the common challenges with embedded evaluation. These strategies, or best practices, for successful embedded evaluation include:

- **Consider the approach:** Consider whether the offering is suitable for embedded evaluation as embedded evaluation may not be the best evaluation approach for every circumstance.
- **Create a shared understanding:** Ensure utility, implementation contractor, and evaluation teams have a shared understanding of embedded evaluation.
- **Shift perspectives:** Adapt perspective from that of a researcher or evaluator to that of a consultant coming in to support successful program implementation.
- **Build flexible scopes and budgets:** Developing flexible scopes of work and budgets to allow for changing evaluation needs.
- **Monitor data collection:** Set the program and evaluation up for success by working with the program team to establish and implement data collection processes.
- **Collaborate regularly:** Collaborate through recurring meetings including the utility staff, IC staff, and evaluation vendors.
- **Rethink reporting:** Consider how and when findings and recommendations will be shared.

Methodology

This paper draws from five case studies of embedded evaluations including two pilots in the west, a statewide program in the Northeast, a pilot in the Northwest, and one new offering introduced by a utility in the Southeast. Drawing from our evaluation team’s experiences with these embedded evaluations and others, we explore strategies to maximize results from embedded evaluation methods and how to respond to or avoid some of the common challenges with embedded evaluation. Because embedded evaluations often provide formative feedback as pilots and programs are first launching addressing early challenges, we keep the utilities and programs for these case studies confidential.

Table 1. Embedded Evaluation Case Studies

Client Region	Program Type	Program Description	Evaluation Type	Evaluation Period
West	Pilot	Income-Qualified Home Retrofit	Process & Impact	2022 - present
West	Pilot	Income-qualified Building Electrification	Process & Impact	2022 - present
Northeast	Program	Implementation of Language Access Plan	Process	2024 - present
Southeast	Pilot	Income-Qualified Whole Home Retrofit, with a Pay-as-you-Save model	Process & Impact	2020 – 2024
Northwest	Pilot	IE Dual Fuel Heat Pump	Process	2024 - 2025

Results

Below, we provide highlights from recent embedded evaluations that helped inform the strategies provided for successful embedded evaluation. For each case study, we describe the offering and point to how our experience connects with the conclusions of this paper, highlighting the best practices used in the lessons learned and key benefits of embedded evaluation to each of the evaluations.

Case Study 1 – Western Income-Qualified Home Retrofit Pilot

An electric utility and gas utility located in a western state jointly designed an income-qualified pilot aimed at maximizing savings for income-qualified households through comprehensive home retrofits. Customer participation begins with an energy assessment to determine the savings potential of the home and qualifying customers can receive comprehensive home retrofits at no cost. Customers must install all recommended measures. One key aspect of this pilot is testing the effectiveness of a customer targeting approach whereby ILLUME as the evaluator worked with the utilities to identify customers most likely to benefit from the pilot based on criteria including energy usage, high bills, rate, and geographic location. Only the identified customers are eligible to participate in the pilot and pilot outreach is limited to those eligible customers.

Table 2. Western Income-Qualified Home Retrofit Pilot

Utility Points of Contact	Utility EM&V Staff & Program Managers
<p>ILLUME Involvement</p>	<p>The utilities designed the pilot with an embedded evaluation in mind and contracted ILLUME as the evaluator shortly after contracting the implementer. At that time, the utility and the implementer were building the infrastructure for the pilot, including establishing data collection forms and processes, and preparing to begin marketing and outreach which allowed ILLUME to establish evaluation data collection needs at the outset and review marketing and outreach materials. ILLUME has continued supporting the pilot as embedded evaluators through regular meetings with utility and implementation contractor (IC) teams, adapting to changing program needs and identifying research tasks to support program delivery along the way. The final evaluation tasks and reporting will occur in 2026.</p>
<p>Key Challenges for Embedded Evaluation</p>	<ul style="list-style-type: none"> • Slow Recruitment: Recruitment for the pilot was much slower and lower than anticipated, partially because of the poor quality of contact information in the utility customer databases and the targeting approach, which limited outreach to customers pre-determined to be eligible. • Lower than anticipated enrollment: ILLUME built the initial process and impact evaluation plans using the program’s participation estimates which assumed very high participation rates among eligible customers and 100% conversion rates from assessment to measure installation. • Data Availability and Access: Despite receiving clearly documented program data collection requirements before outreach began and agreeing to providing the specified data, the implementer did not collect everything specified and to date has been unable to share an accurate data extract with the evaluation team on the cadence requested. • Staff Turnover: Staff turnover in multiple positions at both the utilities and the IC created many challenges, but most notably resulted in confusion over our role as an embedded evaluator and some resistance to supporting the evaluation.

Resolution/Lessons Learned	<ul style="list-style-type: none"> • The evaluation team worked with the utility team to expand eligibility requirements to a larger pool of customers. (Best Practices: Collaborate Regularly, Shift Perspectives) • The evaluation team adjusted both the process and impact evaluation plans to reflect current program conditions and needs. On the process side, we switched from a robust rolling survey approach based on initial participation estimates to participant interviews to gain more detailed insight into the participation experience and prioritized nonparticipant research to understand awareness and barriers to participation among eligible customers. Because initial participation was too low for the initially planned billing analysis tasks, the team undertook a participant characterization to identify characteristics of participants to support program targeting. (Best Practices: Build flexible Scopes and Budgets) • The evaluation team reviewed each implementation tracking data extract for accuracy (e.g., duplicates, missing fields, changes in structure or values) and provided feedback to the utility and IC teams regarding any errors or unexpected changes. (Best Practice: Monitor data collection) • The evaluation team educated new utility staff to orient them to the project and the evaluation approach. (Best Practices: Shift Perspectives, Create a Shared Understanding)
-----------------------------------	--

Case Study 2 – Western Income-Qualified Building Electrification Pilot

A western electric utility designed a pilot to field test an analytical approach to model and target high priority residential customers (reflects those who are most likely to benefit and those who may be underserved by traditional program designs) for potential electrification retrofits a. The utility collaborates with relevant agencies and programs to maximize customer benefits. High level objectives of the pilot are to reduce a customer’s overall energy costs and greenhouse gas emissions by replacing natural gas or propane fueled space and water heaters with electric options and to increase income-qualified customers’ knowledge of and satisfaction with clean energy options provided through electrification.

Table 3. Western Income-Qualified Building Electrification Pilot

Utility Points of Contact	Utility EM&V Staff & Program Manager
ILLUME Involvement	ILLUME joined the pilot as evaluators before recruitment began, as the utility and the implementer were building the infrastructure for the pilot, which enabled ILLUME to specify data collection needs while the implementer was establishing data collection forms and processes. The evaluation objectives include assessing energy and bill impacts as well as pilot processes, market potential, and customer experience. ILLUME has continued supporting the pilot as embedded evaluators through its evolution from initial program launch to its present form, providing timely evaluation research activities along the way. The final evaluation tasks and reporting will occur in 2026.
Key Challenges for Embedded Evaluation	<ul style="list-style-type: none"> • Slow Recruitment: Recruitment for the pilot was much slower and lower than anticipated, resulting in: <ul style="list-style-type: none"> ○ Missing proposed savings targets ○ Wasted administrative budget across all parties ○ Fewer opportunities for mid-flight post-upgrade evaluation activities to support the pilot’s evolution

	<ul style="list-style-type: none"> ○ Eventually, reworking the recruitment and pilot design, which created a need to adjust evaluation plans (and cost money) ● Evolving Program Design: Throughout the pilot, the implementers and program team have modified its design. The evolution has included a variety of changes that require the evaluation team to adapt research plans and instruments and include attempts to increase uptake as well as to avoid customer confusion and inter-program competition from similar offerings. ● Data Availability and Access: Despite early involvement, the implementation and evaluation team encountered hurdles in accessing data needed for the implementation and evaluation, including access to bills for initial bill screening to establish eligibility and billing data from the gas utility for savings impacts.
<p style="text-align: center;">Resolution/Lessons Learned</p>	<ul style="list-style-type: none"> ● To address pilot recruitment challenges, the evaluation team conducted their nonparticipant and implementer field team research sooner than planned to better understand why nonparticipants chose not to participate. As a result of this research, the implementer pursued alternate paths for pilot recruitment and restructured incentives for recruiters to encourage more up-front effort on recruiting participants. This research helped the implementation team successfully boost participation and shift data collection activities to reflect the pilot changes, which could have jeopardized the evaluation. (Best Practices: Build Flexible Scopes and Budget, Rethink Reporting, Shift Perspectives) ● To address the confusion caused by overlapping offerings, the implementation team leveraged a screening tool to identify the most appropriate offering for a customer by estimating bill savings. As embedded evaluators, we coordinated with the program team on their program design changes to navigate the overlapping offerings, which led to two key outcomes: (1) we ensured the adjustments still enabled evaluation and that our research tools, such as surveys, could accurately capture feedback on nuances of the pilot design, and (2) we shifted evaluation resources towards better understanding the screening tool as it became a key element for deciding what customers can receive. Our review of this tool has and will continue to lead to improvements to the screening tool over time, which will in turn improve the customer experience and pilot participation. (Best Practices: Build Flexible Scopes and Budgets, Collaborate Regularly) ● To overcome challenges accessing billing data, the ILLUME team worked with the utility to assess the feasibility of different options for collecting the necessary data. Ultimately, the utility was able to facilitate the data sharing needed with the gas utility, and the added timeline and alternate options ensured the success of the evaluation. (Best Practices: Build Flexible Scopes and Budgets)

Case Study 3 – Northeastern Language Access Plan (LAP) Implementation

Improving language access and increasing equitable participation and access to state-wide energy efficiency programs has been a key aim for the statewide program in one Northeastern state over the past

two program cycles. In the most recent program cycle, the state commissioned the development of a Language Access Plan (LAP) that is being implemented during the next term. The LAP seeks to ensure language access across five programs offered through the statewide portfolio, including Market Rate and Income Eligible residential programs and small business programs. Understanding that there could be research needs and value to having ongoing evaluation support, the Program Administrators established an embedded evaluation for the LAP implementation.

Table 4. Northeastern Language Access Plan

Program Administrator Points of Contact	Program Administrator Implementation, marketing, and EM&V teams
ILLUME Involvement	ILLUME, as part of the residential program evaluation team, had previously conducted research with customers speaking languages other than English (LOTE), including developing a program journey map and identifying barriers for LOTE customers moving through the market rate energy efficiency program, one of several offerings within the broader statewide portfolio. As the Program Administrators launched their LAP implementation across programs, including market rate and Income Eligible residential and small business energy efficiency programs, they invited the ILLUME team to support through an embedded evaluation. This included attending regular implementation meetings, conducting secondary and benchmarking research to support the milestones established in the LAP, and, in the future, conducting fast feedback research to understand how effective changes have been in streamlining the participation experience for LOTE customers.
Key Challenges for Embedded Evaluation	<ul style="list-style-type: none"> • Negative Perceptions of Evaluation: Trust building with implementation teams who previously had negative experiences with evaluators and perceived them as “telling us everything we’re doing wrong” • Evolving Program Design: Planning and startup of evaluation alongside planning and startup of implementation • Unpredictable Implementation Timelines: Implementation vs. evaluation cadence – i.e. the implementation team would sometimes raise questions in meetings that we, as evaluators, could address through a study, but they had a much shorter time-frame for their concern and would make a decision to move forward within the space of the meeting, vs. running a study and find out a response in 6 or 12 or 18 months.
Resolution/Lessons Learned	<ul style="list-style-type: none"> • Integration on regular bi-weekly calls helped to build trust and provided the evaluation team with better context for implementation questions and progress. (Best Practice: Collaborate Regularly) • Evaluation team adapted to a different speed of work and expectations from implementation team, aiming to provide quicker, incremental feedback. The team achieved this by developing new internal processes to accommodate a quicker cadence of client deliverables and expectation setting about those client deliverables. Rather than holding findings until a final deliverable, we provided interim feedback and iterated on these initial versions to develop final work products. (Best Practices: Shift perspectives, create a shared understanding, build flexible scopes and budgets)

Case Study 4 – Southeast Income-Qualified Whole Home Retrofit Pilot

Based on interest from Commission staff in the southeast, the utility team explored a pilot for income-eligible customers that leveraged a ‘pay-as-you-save’ model, where qualifying customers would have a whole home energy efficiency retrofit, including weatherization and potentially new HVAC equipment installed at no cost. The utility covered the upfront cost and the customer paid off a discounted amount over their bills for a period of 7 years. This was the first time that the implementer had delivered an income-eligible version of this program, where no upfront co-pay was required of customers, and that factor ended up being significant in terms of screening customers.

Table 5. Southeastern Income-Qualified Whole Home Retrofit Pilot

Utility Points of Contact	Utility pilot and R&D team
ILLUME Involvement	ILLUME joined the pilot as the evaluator very early in pilot implementation and was involved in the sample design. The ILLUME team evaluated the pilot’s processes, including its ability to achieve the recruitment and savings goals and the participant qualification process; customer experience, including customer’s experiences with the first contact with the pilot, enrollment, installation, and upgrades made; and impacts, including energy and bill savings. The team provided recommendations for the client team as they considered moving this into a full program.
Key Challenges for Embedded Evaluation	<ul style="list-style-type: none"> • Slow Recruitment: Very slow program uptake lengthened the recruitment period from the anticipated 3 months to 18 months. • Lower than Anticipated Enrollment: The implementer invited pre-screened customers to enroll in the pilot via tailored Home Energy Reports, email, and mail outreach letters that pushed customers to the website to complete an application. However, this approach did not result in the projected interest, which created risk for the main evaluation task, which was to understand if the savings would be high enough for the pilot to succeed, which was not possible without adequate participation. • Evolving Pilot Design: As a result of the low participation, the evaluation team adjusted the activities they performed to complete the process and impact analysis, including fielding the survey and interviews at a different cadence than initially planned and adjusting our impact analysis approach to compensate for small numbers of participants.
Resolution/Lessons Learned	<ul style="list-style-type: none"> • The evaluation team was able to use the evaluation budget to support recruitment by interviewing nonparticipants and reviewing marketing materials. The team was also able to incorporate gas savings calculations into the impact analysis. (Best Practice: Build Flexible Scopes and Budgets) • Based on embedded evaluation findings, the implementer and pilot team expanded the geographic eligibility, incorporated gas savings, and lowered the savings threshold for the rebate to increase participation. (Best Practices: Collaborate Regularly, Shift Perspectives, Rethink Reporting) • We found that the customer economics did not, at that time, make this a viable offering for a low-income program. However, we recommended the utility consider a moderate income offer with a limited co-pay (e.g., \$500). One of the challenges the pilot faced was that few homes met the required savings threshold to ensure that customer bills did not increase once the cost of the equipment was added. However, a limited co-pay might have enabled more customers to meet screening criteria to be eligible for the retrofit.

Case Study 5 – Northwest Income Eligible Dual Fuel Pilot

A natural gas utility in the Northwest explored the opportunity of ground-source heat pumps through a dual-fuel pilot in low-income homes.

Table 6. Northwest Income Eligible Dual-Fuel Pilot

Utility Points of Contact	Utility Innovation team and low-income program team.
ILLUME Involvement	ILLUME was invited to provide qualitative evaluation support throughout the pilot. This included customer interviews and on-site observations during equipment installations and interviews with the implementation team and the community action agency (CAP) who was partnering with the utility in supporting the pilot. The team conducted interviews with customers shortly after installations in the summer and again several months later after heating season had finished. This enabled the team to gather primary feedback on both heating and cooling experience with the new equipment.
Key Challenges for Embedded Evaluation	<ul style="list-style-type: none"> • Unpredictable Implementation Timeline: The installation timeline was difficult to set in advance due to unanticipated issues that emerged when drilling. Although the overall installation timeline was on schedule, individual customers did not receive accurate estimated installation dates in advance of their installations. This was a pain point for customers and affected the evaluation as we tried to speak with people shortly after their equipment installation was completed.
Resolution/Lessons learned	<ul style="list-style-type: none"> • Our team was able to remain flexible and adapt to the schedule changes as needed. (Best Practice: Build Flexible Scopes and Budgets) • The resulting report focused on sharing insights from the pilot customers as well as identifying considerations when scaling into a full program. (Best Practices: Collaborate Regularly, Shift Perspectives, Rethink Reporting) <ul style="list-style-type: none"> ○ For example, we were able to provide recommendations for how to address pain points in a larger pilot or program context and highlighted elements that were not an issue in this pilot, but likely would be a challenge to plan around in a full program. ▪ The housing agency/CAP agency conducted the recruitment. When we spoke with customers many of them emphasized their high level of trust in the housing agency and noted that they were eager to participate because of their previous positive experience with the housing agency and their staff. These personal connections would not be scalable in a larger program.

Conclusions

Based on our experiences with the embedded evaluations described above, as well as others, our team documents the value of embedded evaluation, key challenges for embedded evaluation and presents key information and strategies needed to serve as an embedded evaluator and navigate the challenges and mitigate the risks that can come with trying to provide meaningful results when they can be most impactful through embedded evaluation.

Highlights of Value of Embedded Evaluation

The following highlights exemplify the value embedded evaluation can provide to the pilots and programs as shown in the above case studies.

- Supported data-driven program design changes.
- Provided early insights into successes and challenges, ensuring programs have the latest information to inform program strategy and related filings.
- Provided adequate time to resolve data access and availability issues to support evaluation.
- Enabled a successful evaluation by adjusting approach as needed to ensure that research objectives could be met even as pilot design or recruitment strategies changed.
- Informed early decisions on how or if to proceed with a pilot program.

Key Challenges for Embedded Evaluation

The authors provide a consolidated list of key challenges related to embedded evaluation experienced in these case studies:

- Slow Recruitment
- Lower than anticipated enrollment
- Evolving program or pilot designs
- Data availability and access
- Staff turnover
- Negative perceptions of evaluation
- Unpredictable implementation timelines

Best Practices for Embedded Evaluation

Below, the authors elaborate on the following key strategies we believe reflect best practices for conducting embedded evaluation from our experience with these case studies.

Consider the Approach: Consider whether the offering is suitable for embedded evaluation as embedded evaluation may not be the best evaluation approach for every circumstance.

We hear a lot about embedded evaluation, and while there are many benefits of the approach, it may not always be the best approach for the evaluation at hand.

Embedded evaluation is primarily used to explore and inform new offerings and implementation strategies (Massachusetts Program Administrators 2024). An embedded evaluation approach can be used on its own or within a comprehensive evaluation structure. When used in support of a comprehensive evaluation plan, the embedded evaluation may focus on an evaluability assessment or program theory for use in future evaluations and/or findings from embedded evaluation activities may provide early insight and context for a comprehensive evaluation – benefits that often align with newer or more innovative offerings, technologies, or program designs. While achieving desired confidence and precision targets is not generally a focus of embedded evaluation tasks, recurring embedded evaluation activities can achieve desired confidence and precision targets over time which could also support a comprehensive evaluation.

A more traditional, retrospective, or comprehensive evaluation may be the right approach for established programs, and for more data driven programs. In many cases, traditional evaluations offer more opportunities for streamlining evaluation costs. The evaluation team may want to consider what the utility and program implementation teams will need to know that cannot already be answered in real time.

Create a shared Understanding: Ensure utility, implementation contractor, and evaluation teams have a shared understanding of embedded evaluation.

As we experienced in the Western Income-Qualified Retrofit Pilot, not having a shared understanding of the role of the evaluator in an embedded evaluation among all parties involved can lead to frustration, confusion, and even resistance to supporting evaluation activities. It is important for evaluators to recognize that they may be working with utility and/or IC staff who have not been involved in any evaluation before, or only with comprehensive evaluations. Being involved earlier also means that evaluators might be working more closely with utility program managers and IC staff throughout the evaluation and not just at the beginning for a program staff interview and at the end during reporting.

Evaluators should take the time to explain the approach and align on roles and responsibilities so that everyone involved has a mutual understanding of the approach and how it is intended to work to maximize the effectiveness. Evaluators should also keep in mind that it may also take time for staff to fully understand the benefits an embedded evaluation can bring to their offering.

Evaluators should remember to revisit these steps with new team members. As we learned in the Western Income-Qualified Home Retrofit Pilot, knowledge about the approach does not necessarily transfer to new team members and staff turnover can result in additional challenges if new team members are not onboarded to the approach.

Additionally, staff (particularly those who have been involved in comprehensive evaluations) may come in with misconceptions about what embedded evaluation means. For instance, in the Western Whole Home Retrofit Pilot, staff involved after the initial contact initially believed that the only purpose of an embedded evaluation was to establish data collection processes and requests at the program onset so that the program can be evaluated successfully at the conclusion of the program period. This kind of thinking can prevent program staff from fully benefiting from the value of embedded evaluations if not addressed.

Shift Perspectives: Adapt perspective from that of a researcher or evaluator to that of a consultant coming in to support successful program implementation.

Just like utility and IC staff new to embedded evaluation, or evaluation in general, may need to learn about the approach, evaluators newer to the approach also need to learn to adjust their thinking from a traditional program evaluation in order to maximize the success of an embedded evaluation. For example, evaluators might need to get more comfortable taking a more consultative approach and “going with the flow” to some extent - developing research questions and activities to address the changing needs of the program rather than sticking to a rigid evaluation plan designed around meeting confidence and precision targets. And in some cases, embedded evaluations create opportunities for evaluators to weigh in on pilot or program decisions. While valuable for the overall evaluation, these opportunities also create risks for evaluators, where missteps are possible. These risks may have to be mitigated by documenting the reasoning for decisions being made and may require evaluators to keep higher level team members involved throughout the evaluation.

As illustrated in the Northeastern Language Access Plan case study, our team had to adjust their thinking and the way they worked with the implementation team to better support program implementation. By adapting to a faster pace of work, the team provided more frequent, incremental feedback the implementation team could act on right away. This change in perspective is also illustrated in the Western and Southeastern case studies where our team had to pivot from initial evaluation plans,

developed from the perspective of an evaluator, when program participation did not support the planned activities. Our teams had to start thinking less about assessing the program's success and more about how to help the program achieve success to develop revised evaluation plans with the utility and implementation teams.

Build Flexible Scopes and Budgets: Developing flexible scopes of work and budgets to allow for changing evaluation needs.

One of the benefits of embedded evaluation is the ability of the evaluation team to investigate the most pressing research questions at the time to support successful program implementation. While this can sometimes be accomplished through discrete research tasks planned in advance like recurring customer surveys, it often requires more flexibility and the ability to determine the appropriate research activity to address the current program needs as they arise. For this reason, it is critical to design flexibility in scope and budget into the evaluation planning from the beginning. As noted in the 2025 – 2027 Massachusetts Statewide Energy Efficiency Draft Strategic Evaluation Plan, scopes or work plans will need to be more open-ended than typical, and budgets presented within a range to accommodate this flexibility (Massachusetts Program Administrators 2024).

One common theme from the case studies was lower than anticipated participation, or slow program ramp up. This rendered embedded evaluation plans designed using anticipated participation levels and timelines before the program launched useless and raised new, more important research questions. Instead of taking the time to develop detailed evaluation plans before the programs launched, these evaluations may have been better served by more flexible plans that identified groups of interest but not necessarily the research method, or initial plans that specified a cadence for research activities and/or budget range with more detailed scoping taking place as research needs are identified.

Additionally, as part of any scoping or planning documents, evaluators should include guidance for all parties on expectations for and importance of providing timely response to data requests and reviews of research instruments and/or reporting deliverables like PowerPoint presentations and memos.

Monitor Data Collection: Set the program and evaluation up for success by working with the program team to establish and implement data collection processes.

Any evaluation depends on data availability and quality. One big advantage of embedded evaluation is the ability for the evaluation team to be involved in defining data collection needs at program launch. Embedded evaluation also provides an opportunity to identify and correct data collection issues earlier in the program than a traditional retrospective evaluation. However, this does not make embedded evaluation immune to implementation data collection challenges.

In multiple cases, our team encountered instances where the implementer failed to collect the specified data and/or could not extract it from its systems and provide it to the evaluation team on a regular basis. However, having regular meetings with the utility and IC program teams enabled us as the evaluators to address the situation and work with the teams to improve quality data collection.

Collaborate Regularly: Collaborate through recurring meetings including the utility staff, IC staff, and evaluation vendors.

As we have found in our experience, and as noted in the 2025 – 2027 Massachusetts Statewide Energy Efficiency Draft Strategic Evaluation Plan, a successful embedded evaluation requires frequent and close collaboration between evaluation and implementation. With that in mind, the evaluator should work with the utility staff to establish a regular meeting involving all three parties. These meetings provide an opportunity for the evaluator to hear about any program challenges firsthand and proactively work with the program team to identify and prioritize research needs to support program implementation and to ensure evaluability while program designs evolve. They also provide the evaluation team opportunities to share high-level evaluation updates and findings as evaluation tasks are completed and to build trust

and integrate within the larger program team which leads to a more successful embedded evaluation as our team found in the Northeastern LAP evaluation.

Rethink Reporting: Consider how and when findings and recommendations will be shared.

Successful embedded evaluation may require several changes from how evaluators typically present findings from comprehensive evaluations, as detailed in the following paragraphs.

Establish expectations for action. Misalignment of expectations regarding how and when evaluation findings and recommendations will be provided, and acted on, can lead to frustration on both sides. It is important to establish clear expectations for all up front. Utility clients who have not yet established a continuous feedback loop with an implementer may require additional support in communicating suggestions to the implementer and following up on progress toward completion as we found in several of the case studies.

Make recommendations actionable. While good recommendations resulting from any evaluation need to be actionable, this is especially important for embedded evaluations and there are several aspects of making recommendations actionable to consider. To support real-time program adjustments and continuous improvement, recommendations from embedded evaluation activities need to be:

- Shared quickly. Consider providing interim updates through existing meetings and documenting findings in short memos or PowerPoint presentations rather than formal reports designed to be published.
- Reflective of program design or outreach elements the program administrator and/or implementation contractor can make changes.
- Limited to those that are most critical for program delivery – current and future. As demonstrated in the Northwest evaluation, remember to think about scale when evaluating a pilot program – some elements of program design that work well or challenges that are manageable on a small scale may not be sustainable when the program is scaled up.
- Prioritized for immediate attention and longer-term consideration if the number of recommendations is great.

Establish expectations for timing and focus of report reviews. Utility and implementer reviews should be focused on understanding and refining findings and recommendations to maximize the usefulness to the program team. This may be different for some clients who are used to reviewing and providing feedback to prepare for a report to be filed and made public.

Consider which deliverables (if any) should be made public. Further, because findings from these activities tend to be more formative in nature and identify program improvement opportunities when the offering may still be in development, the teams should consider whether any of the deliverables should be made public on a case-by-case basis.

Reference

Massachusetts Program Administrators. 2024. 2025-2027 Three-Year Plan Exhibit 1, Appendix S, 2025 – 2027 Massachusetts Statewide Energy Efficiency Draft Strategic Evaluation Plan, <https://ma-eeac.org/wp-content/uploads/Appendix-S-Strategic-Evaluation-Plan.pdf>