The Program Theory and Metrics Process

Jeff Erickson, PA Government Services Inc., Middleton, WI
Jennifer Fagan, KEMA-XENERGY, Madison, WI
Oscar Bloch, Wisconsin Division of Energy, Madison, WI

Abstract

Wisconsin is approaching the end of its second year of implementing statewide energy efficiency programs using Public Benefits funds. The programs must address a variety of legislative goals including energy efficiency, system reliability, environmental protection, and rural economic development. At the same time, the program implementers must establish a portfolio of programs that balance market transformation and resource acquisition policy objectives. Under direction from the state, the contractors were instructed to take a theory-based approach to program and evaluation design. The goal was to develop metrics that ensure that program motivations and accomplishments are aligned with the broader policy objectives.

Through the process undertaken in Wisconsin and through research on other program theory development efforts, we have identified a range of issues that should be considered in creating program theories and corresponding contractual metrics for program implementation. This paper will begin by defining criteria for good program metrics to set the stage, then it will present several models that describe approaches to defining program theories and metrics. Next, we discuss a wide range of implementation related issues that influence the choice of the approach to be taken and present information on specific issues that were encountered in Wisconsin and how they were addressed. We conclude by summarizing the pros and cons of each approach and describing what we have learned from our Wisconsin experiences that may help to guide the decisions of others embarking down this path.

Introduction

Wisconsin is approaching the end of its second year of implementing statewide energy efficiency programs using Public Benefits funds. The programs are being overseen by the state government but are being implemented and evaluated by private firms. The programs must address a variety of legislative goals including energy efficiency, system reliability, environmental protection, and rural economic development. At the same time, the program implementers must establish a portfolio of programs that balance market transformation and resource acquisition policy objectives. With such a range of issues to address, the process of designing effective and responsive programs and evaluations has proven to be challenging.

Under direction from the state, the contractors were instructed to take a theory-based approach to program and evaluation design. Program designers, state sponsors, and evaluation staff worked together extensively to attempt to design program theories and metrics based on those theories that could be incorporated in implementation contracts. (Program theory, in this context, refers to documenting the detailed chain of events that ultimately lead to energy efficiency measure adoption.) Efforts are now underway to modify programs and develop new metrics for the contracts for the third year of implementation. The goal is to develop metrics that ensure that program motivations and accomplishments are aligned with the broader policy objectives.

Through the process undertaken in Wisconsin and through research on other program theory development efforts, we have identified a range of issues that should be considered in creating program theories and corresponding contractual metrics for program implementation. This paper will begin by

defining criteria for good program metrics to set the stage, then it will present several models that describe approaches to defining program theories and metrics. (For ease of reference, these approaches are referred to as "models" throughout this paper.) Next, we discuss a wide range of implementation related issues that influence the choice of the approach to be taken and present information on specific issues that were encountered in Wisconsin and how they were addressed. We conclude by summarizing the pros and cons of each model and describing what we have learned from our Wisconsin experiences that may help to guide the decisions of others embarking down this path.

The types of programs to which these apply are Public Benefits programs. Public Benefits programs in a given jurisdiction have the following characteristics. They are:

- Administered by state government, utilities, or non-profit organizations
- Funded by a common "public goods" charge
- Designed to fulfill multiple public policy objectives such as resource acquisition energy efficiency, environmental improvement, economic development, and so forth.

Criteria for Program Metrics

There are many factors that have to be balanced when defining metrics for inclusion in a broad statement of goals or in implementation contracts. Those factors include the following.

Timely. Is it likely that the effect will be large enough to be measurable in the contract year? The program theory must provide a plausible story about how program actions will create measurable changes in metric outcomes over the contract period.

Theory-critical. How central to the theory is the metric? How strongly is the metric correlated with ultimate effects? Does the metric form one of the key links between program action and ultimate effect? If metrics do not have a strong role in the program theory, their value for predicting the ultimate effects of the program will be weak and their value as contractual metrics will be low.

Measurable. The chosen metrics should meet several criteria to ensure that they can be reliably and consistently measured. They should be clear and unambiguous. They should present a reasonable possibility that they can be measured with precision at an acceptable cost. They should be amenable to data collection and reporting methods that can be open, visible, and easily verified or double-checked to reduce the likelihood that the program administrator could manipulate the data undetected.

Net or Gross. Should the contract be based on unadjusted data collected and reported by program staff? Or should it be based on evaluation-adjusted results? (Should it be based on gross or net values in impact evaluation terms?) Contracts based on net results will provide motivation to the program to minimize the factors that reduce net savings (e.g., free ridership). However, contracts based on gross results will probably be easier to negotiate and be more predictable.

Possible Models for Development of Program Theory, Metrics and Program Design

In this section, we will present and discuss conceptual models of Public Benefits program development. These models reflect variations in both:

- The number and sequencing of steps in program development and,
- The type of expertise engaged.

Each of these models must produce results that meet the stated policy objectives of the program. In some cases, these policy objectives are established through enabling legislation; in other cases, they are provided by the body that oversees the program operation. Policy objectives include energy

efficiency, market transformation, environmental improvement, economic development, local job creation, and others that are specific to each region.

Phased Implementation Model

The first model discussed is what we call the "Phased Implementation" model. This model, more than any other, utilizes a carefully sequenced step-by-step approach to program development, in which the right kind of expertise is engaged at each step.

As such, this model represents the recommended approach. Not all jurisdictions will have the time or resources to implement this model, but for those that do, it should result in a set of programs that are:

- Research and theory based
- Targeting the right markets
- Offering the correct interventions
- Incenting contractor delivery behavior that is aligned with the program and policy objectives.

Figure 1 below provides a graphical representation of this model:

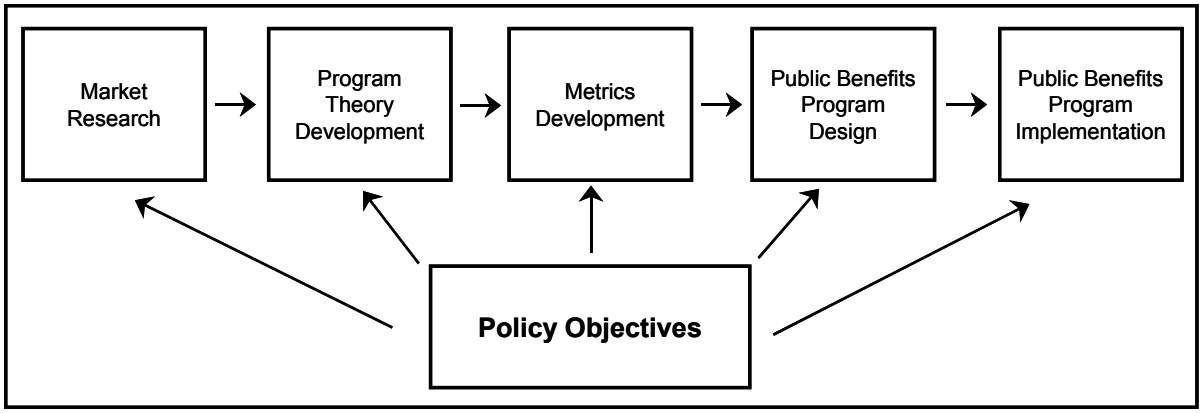

Figure 1. Phased Implementation Model

Another important dimension of this phased approach is the type of expertise that is engaged during each phase. Each phase may use the same or different staffing resources; what is important is that the resources have the right type of expertise for that phase or task.

Market Research is the first phase of program development. Both primary and secondary research is done in this initial step before any further program design activities take place. Staff with *market research expertise* need to be involved in this phase.

Ideally, this research addresses both the demand-side and the supply-side of the markets targeted. An important goal of this research is to provide evidence that can be used to develop program theory and design programs that potentially address both the demand side and the supply side of the market. Another important goal is to provide the basis for measuring change created by the program. The market research ideally will provide the baseline data against which program success is measured. Because of this dual purpose for market research, it is extremely important that the program and evaluation teams work together early in the process to ensure that the market research meets both needs. Because the market research is being done in advance of program design, the evaluation will have to think long and hard about the research approach to maximize the likelihood that data collected will adequately support future performance measurement.

For the demand-side, it is important to research the following areas:

- Awareness of energy efficient measures and programs
- Barriers to implementation of energy efficient measures
- Prior participation in energy efficient programs
- Attributes desired in energy efficient programs

Supply-side research should focus on the structure of the delivery channels and overall markets for each dominant end-use category (e.g., lighting, HVAC, etc). One goal is to try to collect documentation of supply-side market barriers and identify possible solutions.

Program Theory Development is the second phase of this model. The purpose of program theory is to articulate the chain of events that will ultimately lead to adoption of energy efficiency measures, in a format that allows for the identification of intermediate steps in the process and appropriate milestones of progress. The market research performed initially provides much of the basis for the program theory that is developed.

The third phase of **Metrics Development** directly follows. In this phase, a set of meaningful and measurable metrics is established for each program. These metrics represent intermediate and final milestones that can ultimately be measured and serve as the basis for evaluation of contractor performance. The metrics are provided by the well-articulated program theory, which illustrates why it is so important to have completed the earlier research and program theory phases prior to setting metrics. Metrics that are selected without prior research and program theory are more likely to incent the wrong behavior and to lead to other undesirable program consequences.

Those involved in the development of program theory and associated program metrics should ideally have research, evaluation and program design expertise, since the program theory and metrics involve elements of all three.

Fourth is the **Public Benefits Program Design** phase. Program designs articulate the types of markets and end-uses to be targeted, the types of interventions to be offered, and the overall delivery strategy to be used. The bases for the program design developed are the market research, the program theory, and the program metrics developed in the earlier steps. An important function of these earlier phases is to provide evidence and data that can help to inform the program design, so that it will function successfully. Public Benefits program design expertise is needed in this phase. Too often, programs are designed by those with skills in the area of program implementation but with little design experience, which can lead to poorly designed programs (of course the opposite situation can also create problems).

Last is the **Public Benefits Program Implementation** phase. This occurs only after the research, theory, metrics and design phases are completed. This sequencing helps to insure that the programs implemented are based on sound market research, well-informed program theory, and theory-based program designs. Therefore, it is more likely that programs designed in this manner will be successful. This last phase requires those with an implementation background (for example, engineers) to be involved.

Mid-course Correction Model

Often, the organization that is administering the Public Benefits programs is not in a position to implement the phased model. There may be any number of constraints present. In many cases, they find it necessary for them to bypass one or more of the steps described above, and then to revisit the program designs midway through the implementation period. We call this approach the "Mid-course Correction Model" and will describe the basic model plus two variations that we are familiar with. Figure 2 below shows this model graphically.

Under the Mid-course Correction Model, the program skips the research/theory/metrics phases and goes immediately to the design phase followed by implementation. The program design is developed based on prior experience, intuition, and "normal" elements, but without the benefit of the skipped phases. The program theory and metrics are developed after the fact, based on the initial program design.

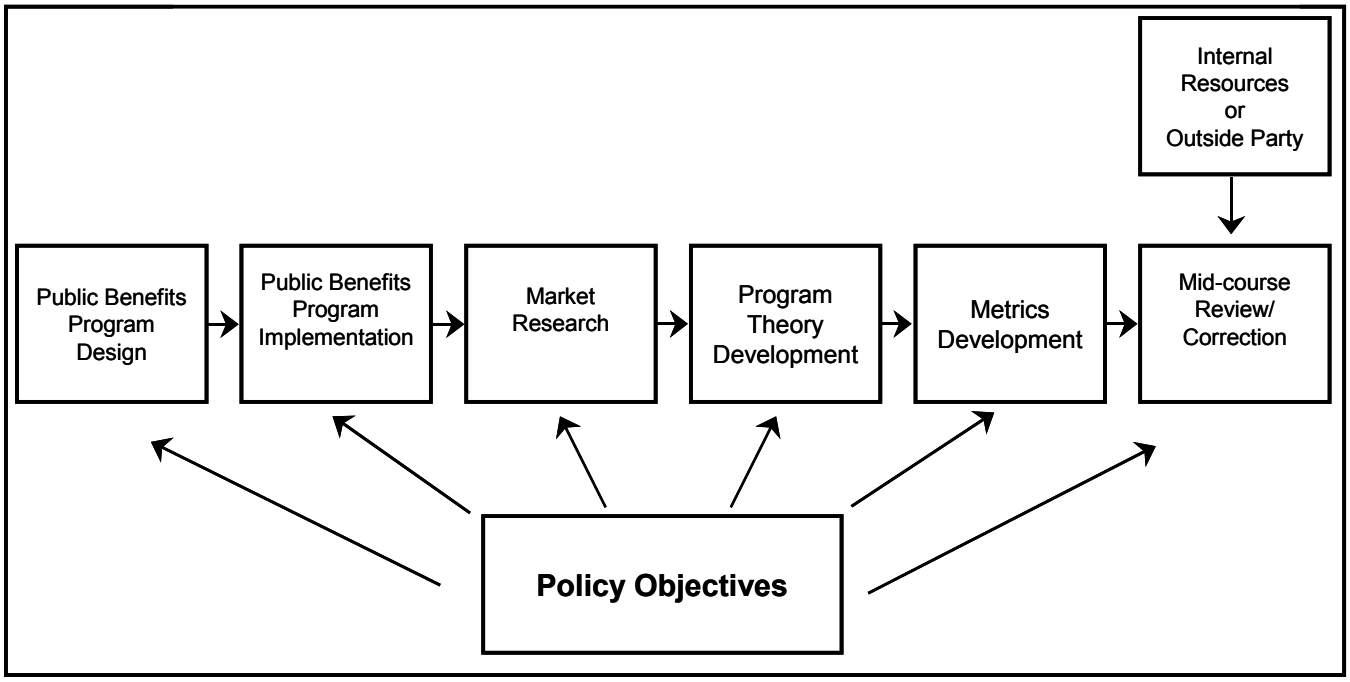

Figure 2. Mid-Course Correction Model

After the initial phase of implementation, the program is found to be missing its target in one or more ways:

- Not meeting stated energy goals
- Not addressing desired markets
- Not inducing desired behavior changes
- Not cost-effective
- Not efficient in terms of its delivery
- Target audience(s) not satisfied.

As a result, the program administrator implements a "mid-course correction" strategy. Experts are brought in to research the problem areas, identify possible solutions, and make recommendations as to how to change the program design in order to address these problems. The experts involved with this activity may be (1) insiders (those already involved with the program); or (2) outsiders (those not involved with the program).

Ideally before recommended design changes are made, the program theory and associated metrics are developed. The revised theory relies on evidence collected since the inception of the program; namely, (1) evaluation findings from primary research activities, such as surveys of program participants and non-participants; (2) additional market research done in connection with the original programs; (3) secondary research and evaluation findings from other jurisdictions implementing the same types of programs; or (4) any combination of the above.

In terms of expertise engaged in the Mid-course Correction activities, there are two variants. Under the first, the program relies on experts already involved in program activities in some manner.

These individuals must have the requisite Public Benefits program design and research and evaluation background and must not have vested interests that undermine their ability to be effective and impartial in this role. It is possible and maybe likely that individuals performing ongoing functions in the current program would not be able to be impartial in this role since they would have a vested interest in maintaining the status quo in order to keep their job. Because of this risk, the model recognizes the value in the second variant where the program engages outside experts (those not involved in the current program) to provide theory and design services. Being external to the program, they, hopefully, will avoid the conflict of interest problems.

Implementation Issues

Which model should you choose? How should you decide? What should you consider in making that choice? Six factors come in to play in determining which model can be chosen: Context, timing, expertise, money, procurement parameters, and independence and objectivity, as illustrated in Figure 3. We address each in turn.

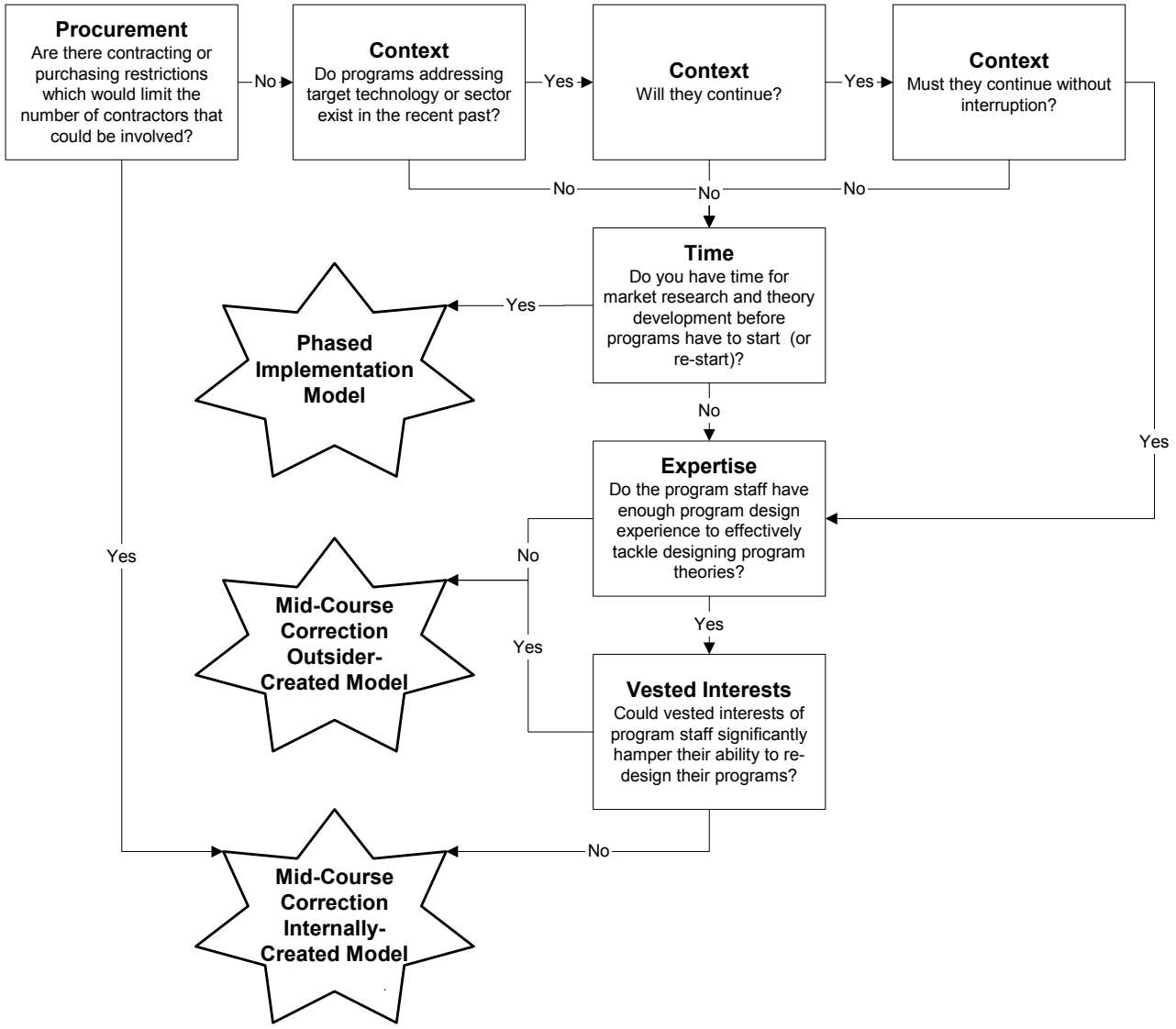

Figure 3. Decision Tree for Choosing a Model

Context

Context has a great influence on which model is appropriate and what information is available to support program theory development. Are programs addressing the target technology and sector currently running or did some exist in the recent past? If so, hopefully research done for or on those programs can provide a wealth of information to inform the program theory and metrics selection process. These data should be thoroughly mined at the beginning of any program theory and metric development effort.

Are the metrics being designed for a program that is already in place? If so, must the programs continue uninterrupted while the theory is defined and metrics developed? If so, then the Phased Implementation Model cannot be used for it assumes that market research and program theory development can proceed unhindered by existing programs. Vested interests from existing programs may influence the range of options considered in developing a program theory and the clarity of the result. The time it takes to do market research and develop a program theory and metrics may be too much to fit in with the schedule of an existing program.

Timing

How much time you have available for research and program design activities can determine which model is appropriate. If programs do not exist that will be modified based on the results of the program theory effort, or if existing programs can be interrupted while the theory work leads to new program designs, and if there is time for all the components, then the Phased Implementation Model is the model of choice for creating effective and logically consistent program designs. The Phased Implementation Model can take longer from beginning to end than the other models, since things are done serially instead of in parallel. However, we would argue it will take less staff time than the other models because the effort can

Wisconsin Case Study #1

Under Wisconsin's Focus on Energy programs, an initial set of activities for some programs were defined without the benefit of significant market research or a well-developed program theory. Subsequently, significant effort was spent designing program theories (using both the variants of the Mid-Course Correction model). However, in the end, program theories were still quite weak for several component programs. Several factors contributed to this problem. First, the request for proposals for Focus on Energy implementation provided significant detail on the types of programs and sectors that should be targeted. It also specified that program theories should be developed. The winning bid came from a team composed of many of the companies who had implemented the pilot program. They proposed to implement the program outlined in the RFP but did not propose any significant modifications. As a result, the statewide programs began as scaled-up versions of the pilot programs with few innovative approaches and no significant effort to develop a program theory that might lead to improved functioning.

The contracts signed with program administrators specified numeric targets for energy savings and cost-effectiveness but their market effects clauses were more vague. State policy was officially in support of programs producing significant market effects but support and emphasis of those aspects of programs from the state was significantly less than the attention paid to the near-term energy savings aspects.

Together these things provided motivation to the program staff for achieving energy savings but not for designing programs well-grounded in theories and likely to produce significant market effects. The motivation for tweaking existing programs to improve their near-term energy savings was strong but the motivation for developing theories that might lead to significant changes in program approach was weak. Since market transformation was officially a goal of Focus on Energy, program implementers strongly defended the market transformation aspects of their programs in discussions about program theories. Yet, they often could not (or would not) successfully describe how some of these activities would lead to changes in markets. Rather than use the development of program theories as a means for culling out ineffective program components – components that did not present promise for supporting changes in markets or directly support resource acquisition – program implementers more often took defensive postures related to all their program activities. They seemed to value highly all of their activities and believe they were necessary, even if they could not articulate how the activities would support program goals and market transformation.

be more efficient. It may take less time to do research, develop a theory, and **then** design a program from scratch than it would to figure out how to fit all the things you are currently doing into a logically consistent and complete theory.

However, if for whatever reason there is not enough time to implement the phases of the Phased Implementation Model in sequence, then one of the other models must be used.

Wisconsin Case Study #2

The Wisconsin Focus on Energy programs attempted to take the Phased Implementation Model approach but were hampered by political and logistic factors that affected the timing. Wisconsin implemented a pilot Focus on Energy program in one utility territory to test the concept of public-benefits energy efficiency programs run through the state. The pilot provided valuable information on program administration, implementation, and evaluation, but it did not provide the kind of market research necessary for effective program design. The state let a market research contract for survey research to establish a baseline, however legislative, political, and bureaucratic barriers prevented it from being completed before programs entered the field. Once the statewide program was approved, political and public pressure became intense for getting programs in the field and operational as quickly as possible. In that atmosphere, an approach requiring sequenced market research and program design before implementation was not feasible.

Expertise

The availability of expertise in several areas can help determine which model is appropriate. Ideally, those involved in the process will have strong knowledge of the characteristics and dynamics of the target market (or at least know good market information when they see it so they can avail themselves of others' research), program theory development, program design, and performance measurement. Deficiencies in any of these areas can significantly hamper the process.

Knowledge of the **characteristics and dynamics of the target market** is critical for identifying the points in a market where judiciously-applied pressure from the program will create desirable improvements in the level of efficiency provided by the market. In ideal circumstances reasonably fresh market research will provide the information needed about the specific market. It is not necessary that the theory developers start out knowing all the details of the specific market, but they should be well versed in the functioning of some markets so they can get up to speed rapidly on the target market using available or newly-collected market research.

Lack of this knowledge can lead to an inefficient program that includes components targeted at unimportant aspects of the market or components that do not have a chance of effecting change because of a size (or power) mismatch between the component and the issue it seeks to address. The program will wrap an ankle when the broken arm is the real problem or it will try to fix a severe cut with a child's Flintstone Band-Aid.

Someone on the team developing program theories and metrics should have some significant **knowledge of uses of program theory and methods of organizing and presenting the theor**y. As we discussed in the introduction, solid program theories have several important characteristics. Involving someone with significant exposure to program theories can help ensure that the criteria of a well thought-through program theory are met.

Needless to say, the program design team should also include individuals with strong experience in **designing programs** – in this context we mean choosing the methods and approaches that will constitute the program, defining how those approaches will be implemented, and tying the approaches to the program theory.

Finally, the program design team should include individuals with expertise in designing performance measurement systems. Their role should be to ensure that the metrics chosen are meaningful and measurable, and that it is clear how data will be collected (and by what organization) to track movement in the metrics. Evaluation staff can typically assist in this role.

If the program staff does not have enough knowledge and experience in energy markets, program theory, and program design to effectively tackle the entire process, then the Mid-course Correction – Outsider Created Model should be used. If they do have enough experience, then the remaining question is about their objectivity, as discussed below.

Independence and objectivity

If program implementation staff is available for participating in the program theory work and program design (or re-design), should they be involved in the process? If so, to what degree? On the one hand, program staff should provide intimate knowledge of their markets and familiarity with some methods that have worked and not worked in the past. They should provide a voice of reason about what is possible to implement. On the other hand, there is the significant possibility that program staff will have vested interests in some methods and market segments and those interests may hinder objective evaluation of program approaches. If this is a significant possibility, the outsider-created model should be seriously considered. An outsider would be more likely to consider each program component on its own merits to see how it fits in the program theory independent of the staff assigned to the task in the past.

Procurement parameters

The final question to ask in choosing a model is: are there contracting or purchasing restrictions which would limit the number or type of contractors that could be involved? The Phased Implementation and Outsider-Created models both assume that new contracts can be established to do research and program design work. If procurement restrictions stand in the way, the only choice may be the Mid-course Correction Internally-Created Model. Also, budget issues, naturally, come in to play – what can you afford to do? Can you afford full-scale, new market research prior to program design? When considering this question, the costs of the effort should be balanced with efficiencies of an effectively targeted and designed program.

Conclusions and Lessons Learned

From a purely process point of view, the Phased Implementation model is the first choice. However, as discussed above, there may be other overriding considerations that lead to selection of one of the Mid-Course Correction models. The following table summarizes the merits and drawbacks of each approach based on these considerations:

Model	Merits	Drawbacks
Phased Implementation	• Correct sequencing of activities • Uses the right expertise in each phase • Will lead to the correct theory, metrics and program design the first time through • Potentially the most efficient model – does not require spending time and resources correcting the program theory, metrics or program design	• May not be time available to allow for research and theory and metrics specification phases to be done first • Budget may not be available to support new, full-scale market research prior to program design • Many circumstances may preclude the use of this model, e.g., current Public Benefits programs cannot be interrupted; procurement process may not allow, etc.

Model	Merits	Drawbacks
Mid-Course Correction Internally-Created	• Enables theory, metrics and design to be revised based on research and evaluation findings to-date and outside expert opinion • No long delays or interruption of program implementation during research, theory and metrics development stages • Easiest of the 3 models to procure required expertise • May be cheaper to implement than outsider-created variant	• Requires multiple iterations of program theory/metrics/design specification. • Those engaged in correction activities may have conflicts of interest with maintaining the status quo • Potential for mid-course "tinkering" rather than replacement or revision of theory/metrics/design. This will make it difficult to attribute cause and effect after revisions are made and may cause other related problems.
Mid-course Correction Outsider-Created	• Enables theory, metrics and design to be re-focused based on research findings • No long delays or interruption of program implementation during research, theory and metrics development stages • Those engaged in correction activities do not have conflicts of interest with maintaining status quo	• Requires multiple iterations of program theory/metrics/design specification. • More cumbersome than insider model to implement since must go through procurement process to obtain outside expertise. • Potential for mid-course "tinkering" rather than replacement or revision of theory/metrics/design. This will make it difficult to attribute cause and effect after revisions are made and may cause other related problems.

Since most administrators of energy efficiency programs are operating in an imperfect world, with many constraints present, there is no "right answer" that will apply to every situation. The choice of models must be made based on the specific context. As the table above demonstrates, each model has its strengths and weaknesses.

From the experience in Wisconsin, we have learned many things:

- Designing programs and developing program theory without benefit of market research and proper sequencing of theory-design steps leads to multiple iterations of program theory specification and program design.
- Program theory, metrics and design developed while the program is running should be revisited mid-way through the implementation period via one of the Mid-Course Correction models. At that point in the program, there should be extensive research data available from evaluation studies and other sources that can help to inform the program theory, metrics and design.
- A mid-course correction strategy can lead to properly specified program theory, metrics, and designs for the back half of the implementation period, especially if program baggage can be discarded and replaced with new program designs that are based in research and theory from the mid-course evaluation activities. However, mid-course corrections may result in major conflicts when there is a need to adjust outdated metrics. Those being held to contract metrics may object to contractual changes, particularly those linked to compensation. Thus, this strategy may not be feasible in such situations.