Impact Evaluation Framework For Technology Deployment Programs

An approach for quantifying retrospective energy savings, clean energy advances, and market effects

Prepared by: John H. Reed, Innovologie LLC ◆ Gretchen Jordan, Sandia National Laboratories ◆ Edward Vine, Lawrence Berkeley National Laboratory

July 2007

U.S. Department of Energy
Energy Efficiency and Renewable Energy
Bringing you a prosperous future where energy is clean, abundant, reliable, and affordable

This page is intentionally left blank.

Acknowledgements

This work was completed for the U.S. Department of Energy (USDOE) Office of Energy Efficiency and Renewable Energy (EERE) by Sandia National Laboratories, Albuquerque, New Mexico, USA, under Contract DE-AC04-94AL8500. Sandia is operated by the Sandia Corporation, a subsidiary of Lockheed Martin Corporation.

John H. Reed (Innovologie LLC), Gretchen Jordan (Sandia National Laboratories), and Edward Vine (Lawrence Berkeley National Laboratory) developed the impact framework and prepared the document. Jeff Dowd (USDOE EERE, Office of Planning, Budget and Analysis) was the DOE Project Manager.

Charles Bailey, Moria Morrissey, and Jeff Riggert of the Innovologie Team assisted in creating this document. The authors also thank Yaw Agyeman of TMS Inc. for his assistance in managing the review process and preparing the document.

Special thanks are due to those who provided comments that were helpful in improving the framework.

EERE Reviewers	External Peer Review Panel	Other Reviewers
Robert Gemmer Faith Lambert	Floyd Barwig Iowa Energy Center	Joel Eisenberg Oak Ridge National Laboratories
	Ken Keating Bonneville Power Administration (retired)	Michaela Martin Oak Ridge National Laboratories
	Lori Megdal (Chairperson) Megdal and Associates	Martin Schweitzer Oak Ridge National Laboratories
	Carol Mulholland PA Consulting	
	Patricia Rogers Royal Melbourne Institute of Technology	
	Polly Shaw California Public Utilities Commission	
	James Winebrake Rochester Institute of Technology	
	James L. Wolf Independent Consultant	

Notice

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

Executive Summary

The goal of the impact evaluation framework is to help program planners, managers, and evaluators of technology deployment programs organize, design, and implement retrospective impact evaluations that:

Toward a framework for developing powerful and meaningful impact evaluations.

- Realistically assess progress toward the key goals of a program.
- Focus on those things that program planners and managers control and/or influence.
- Give credit that is due to program managers for the direct and indirect effects clearly attributable to their programs.
- Provide a theory-based scientifically grounded approach.
- Produce credible evaluations.

The impact evaluation framework is specifically designed to assist energy program managers and evaluators in Federal, state, and local governments and in public entities and institutions that are increasingly accountable for delivering and demonstrating results. The framework provides a series of steps and some templates that evaluation contractors and program managers can use to develop powerful and meaningful impact evaluations to help refine their programs, increase program effectiveness, make the tough decisions to drop ineffective program elements, and develop credible evidence to help communicate the value of the program to stakeholders. The framework emphasizes linking program outputs to short term and long term outcomes (impacts), measuring target audience response to program outputs, designing sound evaluations, and taking credit for *all* of the program effects that are attributable to the program. An additional aim is to help demonstrate the cost effectiveness of programs.

Program managers usually have a good handle on the quantities of products and services they deliver. The more difficult problem is connecting those outputs with outcomes in a way that stakeholders find believable, convincing, and replicable. Case studies and testimonials that document target audience behaviors in response to program outputs help but are often not enough.

The challenge is to build evaluations that are believable, convincing, and replicable.

Those who are skeptical of a program's goals or share the goals but are skeptical of the implementation approach are often quick to counter such case studies and testimonials by suggesting alternative explanations for changes in target audience behaviors and/or pointing to competing examples as a basis for questioning program accomplishments.

A Different Framework

A good set of tools and templates can help focus the evaluation on defining and measuring outcomes.

This impact evaluation framework is quite different from approaches found in other documents or books about impact evaluation. Typically, those publications are technically oriented treatises focusing on measurement and analysis techniques and are helpful when the linkages between the outputs and outcomes and what needs to be measured are known. Here, we attempt to provide managers and their evaluators with specific tools to define the linkages between outputs and outcomes. This will allow application of established principles of social science to more clearly identify what needs to be measured, develop better evaluation designs, and more effectively harness existing data collection activities to obtain needed data.

Identifying the linkages between outputs and outcomes is one of the most critical and difficult problems in program design and evaluation.

We start from the premise that identifying the linkages between outputs and outcomes is one of the most critical and most difficult problems in program design and evaluation. We present tools and strategies for increasing the understanding of partner and target audience responses to a program so that those responses can be measured more precisely and differentiated from the responses of people and organizations not specifically targeted by the program.

The Impact Evaluation Framework

There are seven key steps in the framework.

This broad framework involves a seven step process:

1. Identify scope, objective, and priorities.
2. Select the types of evaluation to be completed.
3. Select the aspects of deployment-induced changes to be evaluated.
4. Identify research questions and metrics.

5. Design the evaluation.
6. Conduct the evaluation.
7. Report and use the results and data.

Each of these steps involves sub-steps.

Domains

EERE programs target four groups of partners or target audiences.

The impact evaluation framework focuses on four groups of partners or target audiences that deployment programs, such as those in the Department of Energy's Office of Energy Efficiency and Renewable Energy, most typically engage: the knowledge community; public entities including government and nongovernmental organizations; business entities such as manufacturers, business, and professional service providers; and end-users. Deployment programs may have intermediate goals or desired outcomes in each of these domains.

1. The knowledge community: Create, advance, and package market and technical knowledge to make energy efficiency more accessible and implementable.
2. Public entities: Change policies, structures, and operations to smooth the advancement of energy efficiency.
3. Manufacturers and other businesses: Create and enhance products, create and align market channels, enhance marketing, and develop installation and support infrastructures.
4. End-users: Adopt, replicate, institutionalize, and enculturate energy efficient technologies.

If the players in these four domains respond to program activities, then the deployment program can meet its ultimate goals of reducing energy use and energy intensity, increasing the use of clean energy, and reducing harmful emissions while enhancing productivity and global security.

The need is to assess whether or not the program outputs successfully result in the audiences doing these things, and if so, to understand whether or not these actions lead to the desired impacts. The purpose of this framework is to make the linkages transparent, measure the results, and attribute the results to deployment activities.

Diffusion of Innovations

The framework is based on a powerful explanation of social change, the diffusion of innovations.

The impact evaluation framework draws upon diffusion of innovations (Rogers, 2003), which is widely used to describe how social change occurs. Diffusion of innovations is based on literally thousands of studies and has been used extensively in health, social services, education, and, to a lesser extent, in the energy field.

Five basic elements of the diffusion of innovations are:

- The diffusion process — Audiences become aware of technologies or practice. They collect and sort information that forms a basis for a decision about a technology or practice. They decide whether or not to adopt. They implement the adoption decision. Then they confirm the decision. When audiences adopt they may subsequently replicate the decision in other settings, continue to use the technology or practice into the future, become a champion, and/or institute organizational and structural changes to sustain use of the technology or practice.
- The sociocultural environment — Change takes place in social, political, cultural, and market contexts that can accelerate or impede the advance of technologies or practices. It is important to understand and sometimes to take action to change these environments.
- Audience characteristics — Individuals and firms accept innovation at different rates. Individuals and firms are often labeled as "innovators," "early adopters," the "early majority," the "late majority," and the "laggards."
- Product characteristics — Products that are widely adopted typically have relative advantages over other products. They are compatible with existing social, cultural, and social systems. They have limited complexity. They can be tried. And, the results of their use are easily observed.
- Communications mechanisms — Information can be spread by broadcast – one to many methods – or by "contagion" or one-to-one contact. One-to-one contact through networks is one of the most powerful ways in which information spreads.

Domain-Specific Templates

The framework provides a set of tools that can be used to structure impact evaluations of deployment programs. A description of the diffusion of innovations is provided in Chapter 2. The theory is then translated to a generic flow diagram (Chapter 4) that describes how change occurs. The diagram is then applied to each of the four domains: the knowledge community, public entities, manufacturers and other businesses, and end-users. The generic diagram, along with the four domain-specific diagrams, provides templates for describing in more detail how change can occur in each of the domains. The diagrams are intended to be suggestive rather than exhaustive.

The framework provides a set of tools or templates to assist users in describing change in each of the four domains.

For example, program managers and evaluators can take the diagram for a program that addresses manufacturers or other businesses, test it against their own programs and experiences, and begin to systematically examine how their own programs work. Some elements of the diagrams may have to be modified and others added, and some (or all) are likely to need more detail. These schematics are intended to help program managers and evaluators:

- Identify intermediate and long term outcomes (impacts) in detail.
- Be more systematic in identifying the elements and linkages that cause program impacts.
- Look at their programs more systematically.
- Drive this process with the best available social science.

Also, a tool is provided to help program managers and evaluators identify important evaluation questions and metrics. The table in Chapter 5 presents a comprehensive set of generic evaluation questions organized around diffusion of innovations concepts. The intent is for users to pattern their own questions and metrics for the appropriate domains after the questions in this table. The intention is not to replicate every question and every metric, but to systematically identify a small subset of questions that are central to the evaluation being performed.

Another tool is designed to assist in identifying evaluation questions and metrics.

Evaluation Design

A good evaluation design ensures that programs get the credit they deserve for their activities.

This document also focuses on the issue of evaluation design, a much neglected topic. The idea is to assist managers and evaluators to identify cost effective evaluation designs that will help them more clearly separate the effects of their program from effects from other sources.

Data Collection Strategies

Routine data collection makes it easier to do timely and sound evaluations.

Impact evaluations often involve reconstructing data after the fact, which is one of the reasons that they can be so costly. This framework emphasizes the potential for using and upgrading existing data collection procedures to support evaluations. Program managers will find that enhanced routine data collection offer opportunities to do more evaluation at less cost and to provide better evaluations.

Examples

In Chapter 8, examples of some of the various concepts presented throughout the document are provided. Wherever possible, examples have been drawn from studies completed for EERE although studies from other areas have been drawn upon as well.

Using the Document

Program managers will want to be generally familiar with the approach outlined in this document. Evaluators will want to have an in-depth understanding of the framework.

In addition to this document there is a brief companion paper (Reed, Jordan, and Vine, 2007). The companion paper 'walks through' an example of the application of the framework to an actual program evaluation and highlights the flexibility in applying the framework at the sub-program, program, or portfolio levels.

Some users will find it useful to read this document from cover to cover while others will want to peruse some sections and seriously study others.

Others may wish to pick and choose some portions of the document such as the description of the framework, the diagrams showing outcomes in the end-user, business, public entity, or knowledge domains, the table providing examples of impact measures, the discussion of evaluation design, or the examples of measured outcomes. Margin notes make it possible to peruse the document and capture the story line or pick up new ideas.

Users who are interested in the diffusion of innovations, but do not want to read Rogers' very substantial text, may find Chapter 2 a very useful entrée to the underlying theory and concepts. Chapter 3 introduces the overall framework and reinforces the diffusion concepts in Chapter 2, providing substantial additional insight into diffusion of innovations.

Table of Contents

Acknowledgements ..... **i**

Executive Summary ..... **iii**

A Different Framework iv
The Impact Evaluation Framework iv
Domains v
Diffusion of Innovations vi
Domain-Specific Templates vii
Evaluation Design viii
Data Collection Strategies viii
Examples viii
Using the Document viii

Table of Contents ..... **xi**

1. An Impact Evaluation Framework for Refining and Demonstrating Program Performance **1-1**

1.1 Good Evaluations Begin with a Few Key Questions 1-3
1.2 A Generic Program Theory for Technology Deployment Activities 1-4
1.3 Why a Framework 1-6
1.4 The Framework — Resource Acquisition and Market Transformation 1-8
1.5 How to Use this Document 1-9
1.6 Linkages between This and Other EERE Reports 1-10

2. Rogers' Diffusion of Innovations **2-1**

2.1 An Overview of Diffusion 2-1
2.2 Diffusion Process Model 2-2
2.3 The Physical, Social, and Cultural Environment (Market Model) 2-4
2.4 Communications Model 2-5
2.5 Characteristics of Adopters (Model of Adopter Types) 2-8
2.6 The Product Model or the Characteristics of the Innovation (Product Characteristics Model) 2-10
2.7 Why Diffusion of Innovations 2-12

3. Framework for Impact Evaluation **3-1**

3.1 Module 1: Identify the Scope, Objectives, and Priorities of the Impact Evaluation . 3-1
3.2 Module 2: Select the Type of Evaluation to Be Completed 3-12
3.3 Module 3: Select the Deployment Changes to Be Evaluated 3-15
3.4 Module 4: Identify Researchable Questions and Metrics 3-25
3.5 Module 5: Study Design 3-25
3.6 Module 6: Conduct the Evaluation Study 3-26
3.7 Module 7: Present the Results and Store the Data 3-27

4. Using Diffusion of Innovations to Define Outcomes and Impacts 4-1

4.1 A High Level Generic Logic Model of Technology Deployment Activities 4-1
4.2 Application of Diffusion of Innovations to Defining Market Effects and Impacts ... 4-6
4.3 Integrating the Adapted Model of Change with the Logic Model 4-11
4.4 Application to the End-user Domain 4-11
4.5 Application to the Business Entities Domain 4-15
4.6 Application to the Public Entities Domain 4-19
4.7 Application to the Knowledge Entities Domain 4-22

5. Framework Questions and Metrics 5-1

6. Developing an Evaluation Design 6-1

6.1 Evaluation Design Options 6-2
6.2 Direct and Indirect Impacts of Deployment Programs 6-2
6.3 The Double Counting Problem and Attribution 6-7
6.4 Evaluation Design Using Randomized Controlled Trials (RCTs) 6-8
6.5 Comparison Group Designs — the Alternative to RCTs 6-11
6.6 Other Comparison Strategies 6-15

7. Design of Data Collection Activities 7-1

7.1 Routine Data Collection 7-1
7.2 An Example of Implementing Routine Data Collection 7-4
7.3 Secondary Data Sources 7-8

8. Reporting Outputs 8-1

8.1 Energy Savings Reporting 8-1
8.2 Market Effects Reporting 8-3

9. Summary 9-1

9.1 The Framework 9-1
9.2 Domains 9-2
9.3 Diffusion of Innovations 9-2
9.4 Domain-Specific Templates 9-3
9.5 Evaluation Questions and Metrics 9-4
9.6 Evaluation Design 9-5
9.7 Data Collection 9-5
9.8 Some Illustrations 9-5

10. Acronyms 10-1

11. References 11-1

12. Index 12-1

1. An Impact Evaluation Framework for Refining and Demonstrating Program Performance

This document is designed to assist technology deployment program managers and their evaluation contractors in developing an impact evaluation. It describes a framework for implementing evaluation studies to estimate retrospective energy savings and the market effects (actions taken by the target audience(s) in response to the program) of deployment and technology delivery activities. The goal of the framework is to help program planners, managers, and evaluators working with the U.S. Department of Energy's (USDOE) Office of Energy Efficiency and Renewable Energy (EERE) organize, design, and implement impact evaluations that:

- Realistically assess progress toward the key goals of a program.
- Focus on those things that program planners and managers control and/or influence.
- Give credit that is due to program managers for the direct and indirect effects clearly attributable to their programs.
- Provide a theory-based scientifically grounded approach.
- Produce credible evaluations.

In implementing evaluations, program managers need to maintain an "arms length" relationship between evaluators and themselves. This creates an extra burden on the part of programs to create transparent and defensible evaluation processes and conduct quality evaluations using independent evaluators.

The framework is intended to help evaluators and program managers — working with policy planners and others — cut through the complexity and uniqueness of deployment programs to create relatively simple characterizations of behavior. These characterizations can be used to understand what is important, to develop appropriate measures, and to implement the measures in

a way that will make clear what programs are and are not doing. The result is intended to be high quality and defensible (unbiased) impact evaluations.

This is a relatively simple and flexible framework that evaluators and program managers can use to structure evaluations. It allows users to:

- Clearly describe program goals.
- Show how program activities and outputs are designed to influence partners and target audiences.
- Describe how partners and target audiences are expected to respond to program activities and outputs.
- Identify relevant metrics that capture what partners and target audiences do in response to the program.
- Utilize appropriate and relevant evaluation designs and methods to measure changes in partner and target audience behaviors.
- Present analyses that demonstrate whether or not those responses result in a program meeting its goals.

Figure 1 provides an overview of the seven modules in the framework. Module one addresses the scope, objectives, and priorities of the evaluation.

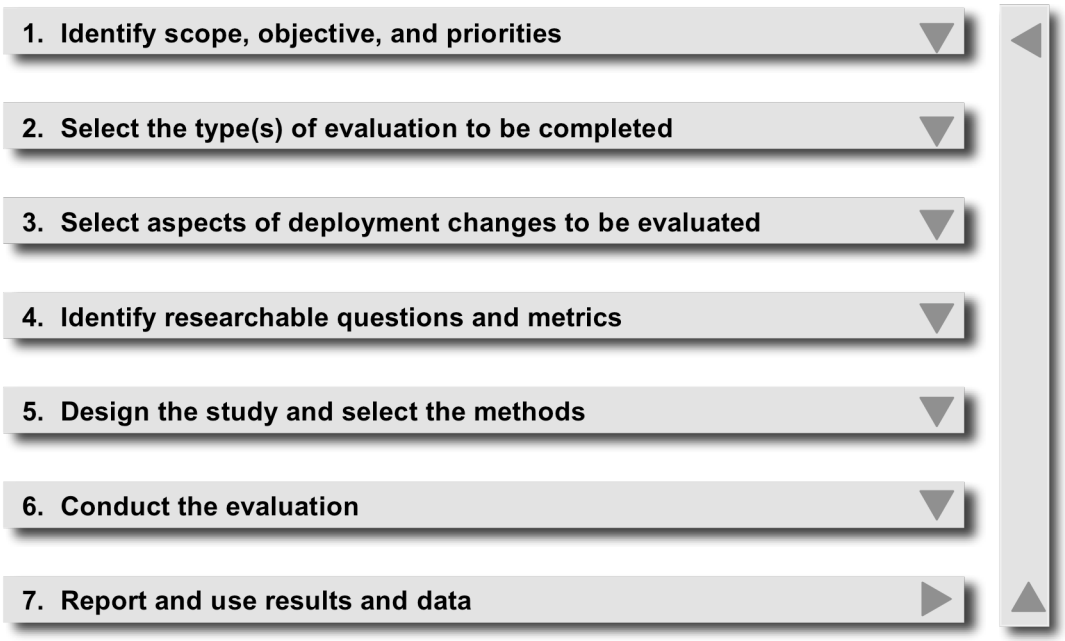


Figure 1. Overview of the Impact Evaluation Framework

Module two is focused on the types of evaluation to be completed. Module three addresses the different aspects of a program that are to be evaluated. Module four focuses on evaluation questions and metrics. Module five addresses the critical issue of study design. Modules six and seven address implementation and reporting. The bar at the side indicates that the results should feedback to subsequent evaluation efforts.

1.1 Good Evaluations Begin with a Few Key Questions

To develop good evaluations that will guide program efforts and clearly demonstrate value to stakeholders, it is important to ask a series of questions.

What are the desired market effects and impacts?

The first question to ask is, "What are the program's desired long term goals (desired impacts)?" For energy program managers, the usual answer is to reduce energy use, to produce more energy from renewable sources, and/or to reduce emissions and pollution released to the environment. For individual programs, the goals may be more specific. For example, the goal might be to reduce energy use in low-income households, in new buildings, or in industrial plants, and/or to design buildings that produce as much energy as they consume.

Who must decide and act: end-users, market actors, policy makers, knowledge workers, or all of the above?

Once the impacts are clearly defined, the next questions are who is to produce the impacts and how are they to be produced. The ultimate answer is that decision makers in households, industrial plants, commercial buildings, transportation organizations, and public facilities must decide to change existing behaviors and use efficient technologies and practices or to install and use renewable technologies.

Does the program need to condition the market so that it supports the technology? Is the program conditioning the market? Does this need to be a part of the evaluation?

However, the decisions and actions of end-users cannot occur in a vacuum. There are a whole series of necessary and sufficient conditions that enable the actions of end-users. For instance, if the objective is to introduce end-users to a new technology, there must be a viable market in which to deliver the technology. Efficient technologies must be manufactured and distributed. Distributors and retailers must stock them and efficient practices

must be available. Professionals such as architects and trades persons must know about them and promote them or decide to use them. There must be people to use, install, and maintain them.

Another set of necessary conditions has to do with the public policies and institutions. Are public policies in place that inhibit or promote a technology or policy? Are public institutions in place to promote technologies and practices? For example, do local codes prevent the use of certain technologies? Does the lack of public or private standards, such as net metering rules, inhibit decision makers from using a technology or practice?

Does the program need to condition the public policy arena? Is the program conditioning the public policy arena? Does this need to be a part of the evaluation?

A last important set of conditions has to do with the infrastructure for knowledge creation and dissemination. Are there organizations or institutions in place to produce knowledge? Is knowledge being created? Is knowledge being packaged in ways that make it usable for others? Has the program attempted to do this and has it been successful?

Does the program create and package knowledge? Does the evaluation need to address the effectiveness of knowledge products and knowledge transfer?

The deployment of a technology or a practice is dependent on an interconnected set of activities across the knowledge, public, business, and end-user domains. The program may only address some domains and other types of organizations, such as nongovernmental organizations, may address others. However, there are likely certain necessary conditions in all domains that need to be addressed for a program to be successful. Thus, it is important to know these necessary conditions.

1.2 A Generic Program Theory for Technology Deployment Activities

Figure 2 provides a generic high level description of how typical deployment programs produce impacts. At the top level, programs analyze and plan; build infrastructure; fund and promote the adoption of new technologies; and review, evaluate, and report. Programs target knowledge workers, public entities, market players, and end-users or some combination of these. Program delivery activities usually aim to cause one of four things:

Deployment programs target four domains:
- *Knowledge*
- *Public entities*
- *Business*
- *End-users*

- Create and package knowledge to make it accessible.

- Condition public policies and institutions to facilitate the delivery of energy efficient and renewable technologies.
- Condition the market to promote energy efficient and clean energy technologies and practices.
- Influence end-users to adopt energy efficient and clean energy technologies and practices.

In other words, programs, in varying degrees, conduct activities to influence audiences in four domains:

- Knowledge
- Public policy and public institutions
- Business
- End-user

The desired outcomes are for people, firms, and organizations in these domains to respond to program activities and outputs and take steps to produce the desired impacts. In general, deployment programs have neither the

Programs cannot reach everyone. Others need to take up the cause.

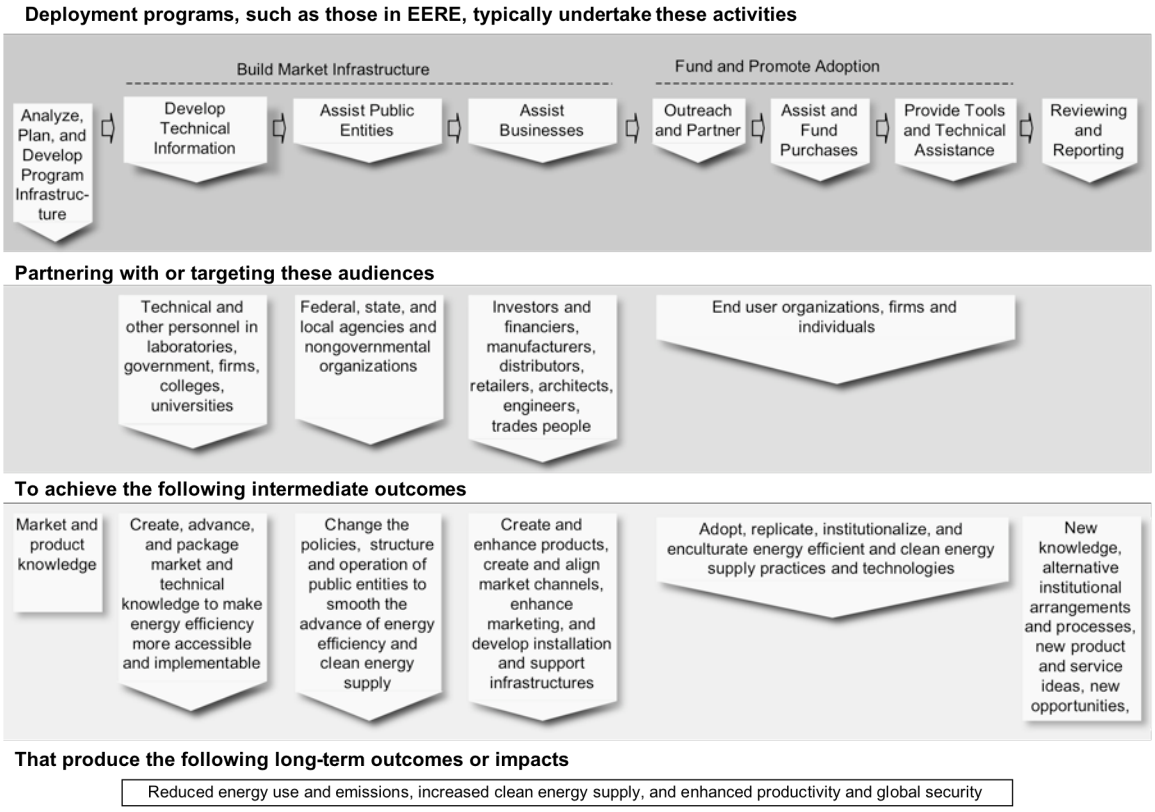


Figure 2. A High Level Program Logic for Technology Deployment Activities

staff nor the resources to continuously stimulate actions in these domains or to touch all possible actors. Thus, it is important to create change among the actors in these domains in such a way that they repeat and sustain their actions in the absence of stimulation from the program.

The actors in the domains are not independent of one another. Although not shown in the diagram, there is feedback within and between domains. For example, actors in the business domain may lack information, so actors in the knowledge domain respond to fill this gap.

Sustained change occurs when target actors replicate their behaviors, their counterparts emulate their behaviors, and both the target actors and their counterparts internalize the new behaviors.

Furthermore, it is important that actors who are not directly involved with programs observe the actions of those who are and learn about and emulate actions. It is through replication, emulation, and internalization that efficient and clean energy technologies and practices become a part of the culture.

It is relatively easy for programs to document activities and the outputs of activities. The Office of Management and Budget (OMB), Congress, and management in Federal agencies such as DOE's EERE want evaluations that go beyond the counts of program outputs. The evaluation must demonstrate that those outputs have triggered actions and reactions on the part of partners and target audiences and, furthermore, that these actions would not have occurred without the program. This puts the onus on program planners, implementers, and evaluators not only to describe the actions of target audiences in these domains, but also to identify and measure change. These measurements must provide clear evidence that, in response to programs, behaviors have changed from a baseline and that other explanations are ruled out.

1.3 Why a Framework

Managers of deployment programs have tended to assume that target audiences will respond to training, audits, financial assistance, and other services. Through episodic observation and intuition, they come to believe that a program's services cause people to act. Managers know what programs have

done and can see the end result. As a result, managers make the "leap of faith" that what the programs have done and the end results are connected.

But, are they connected? Is it possible to demonstrate that what the program has done is linked to an action, which causes a reaction, which causes a reaction, which leads to the end result? For the most part, managers have not described the steps in this process, nor have managers identified the measurements, collected the data, or used the methodologies required to demonstrate linkages.

What programs do

↓

the magic in the middle leads

↓

to impacts.

This is not an easy task. In fact, one might despair somewhat because if one looks at just one domain, say, the business domain, one realizes that there are a very large number of players and a very large number of actions and reactions that occur. Program managers and evaluators realize that they are trying to understand what is happening in not just one domain but two, three, or four apparently unique and complex domains that vary from program to program and have little commonality.

What programs do

↓

cause reactions

↓

which cause reactions

↓

which lead to impacts.

The framework is comprised of two highly integrated parts. The first part of the framework is a seven-step process intended to guide users through basic decisions about the scope, content, design, methods, analysis, and presentation of evaluation results. The second part of the framework is a method designed to facilitate the development of a characterization and understanding of the likely responses of actors to program interventions in the knowledge, public, business and end-user domains.

The framework, based on the diffusion of innovations (Rogers, 2003), is designed to help program managers and evaluators apply one of the most comprehensive theories of social change and to provide a technique to create a generic program logic of the probable actions of partners and target audiences in as many as four domains. Program managers often develop their own theory when designing a program, creating ad-hoc explanations for what occurs in each of domains. This framework offers program managers the opportunity of using a scientifically based description of behavior generically

Take the mystery out of the magic in the middle — create a story for how target audiences might behave.

applied to each of the four domains, with the ability to adapt the generic descriptions to their needs. A program manager who is, for example, attempting to condition a building energy program, can use the description for the business domain logic as a basis for developing a program-relevant logic model of the expected responses for the commercial building market to his or her program.

This impact framework has substantial value. First, it brings to bear the best insights from social science to help managers describe how their programs operate. This framework provides a scientifically based approach for describing behavior that will make it easier to identify and measure program outcomes. Secondly, program managers and evaluators can rapidly tailor the theory to their specific program. This helps to reduce the often laborious process of developing one's own explanations for how the world works. This approach also helps to identify activities, behaviors, and outside influences that may not be a part of a program design or factors that should be taken into account.

This framework also has the potential to help program managers and evaluators organize their thinking and focus on the most important behaviors and on the most important things to measure. This improved focus can result in parsimonious and cost effective designs, methods, data collection, and analysis procedures.

The framework can be used to develop a clear rationale for measures supplied for the Government Performance and Results Act (GPRA) and DOE's performance tracking system (JOULE). Finally, the framework provides a logical basis for evaluation design that will help programs meet the requirements of OMB's Performance Assessment Rating Tool (PART).

1.4 The Framework — Resource Acquisition and Market Transformation

The impact evaluation framework is designed to provide guidance and is a starting place for all deployment program impact studies. The degree to

which the framework is applied depends on the specific nature of the program being developed. Ultimately, evaluations are limited by budgets and time-frames. The framework is intended to help define and sort through evaluation options that will result in more focused and targeted evaluations.

Since the 1990s, there has been a lot of discussion of the differences between market transformation and resource acquisition programs. Market transformation programs seek to change the market -- to ensure that technologies, services, and practices are energy efficient. Resource acquisition programs seek to achieve energy savings in the near term, primarily through the installation of energy efficiency technologies or changes in behavior. For evaluation of resource acquisition programs, some have argued that it is only necessary to assess energy savings from the direct uptake of a particular technology or practice and determine if the effort is cost effective. Little attention is paid to market effects or other outcomes. However, as soon as one begins to adjust energy savings to account for free riders or program spillover, then market effects and other outcomes need to be addressed.

1.5 How to Use this Document

This document is designed to assist in developing an impact evaluation. Users can read it from front to back or go directly to sections of interest. Because the diffusion of innovations' concept provides the underpinnings of this approach, a summary is provided in Chapter 2. This information provides valuable insights into why some programs and program elements work and others do not based on a systematic examination of how social change occurs. Chapter 3 discusses the overall framework step by step, identifies issues and ideas, and provides examples of how to approach an evaluation. Users who want to organize their thinking about an evaluation should find this overview very useful.

Chapter 4 provides templates that can be used to analyze the end-user, business, public entity, and knowledge domains to describe what partners and target audiences do with the outputs of a program. The templates can be used

as aids for thinking more systematically about how a program can ensure desired impacts. Chapter 5 presents a table containing a systematic set of evaluation questions and metrics. By referring to this table, the user can identify and select evaluation questions and metrics and adapt them to his or her program.

Chapter 6 discusses evaluation design and provides practical examples of ways in which programs can design studies to identify linkages among a program, its component parts, and its outputs, outcomes, and impacts. The subsequent chapter discusses data collection strategies and emphasizes the importance of routine data collection. Chapter 8 concludes by showing a number of examples of findings that are drawn from EERE as well as other evaluation studies that illustrate the points made throughout the document. A companion document provides a brief example of the application of the framework.

1.6 Linkages between This and Other EERE Reports

This document complements other reports prepared by EERE:

1. The *EERE Peer Review Guide* (US Department of Energy, 2004) describes steps to plan, design, and implement external peer reviews. Peer review is a form of process evaluation.
2. The *Guide for Managing General Program Evaluation Studies* (Barnes and Jordan, 2006) focuses on the management and use of general program evaluation studies performed by outside experts and contractors. In addition to outcome and impact evaluation, the guide addresses process, cost-benefit evaluation, and market assessment evaluation.
3. *Overview of Evaluation Methods for R&D Programs* (Ruegg and Jordan, 2007) is a booklet that introduces managers to a variety of methods for evaluating R&D programs.

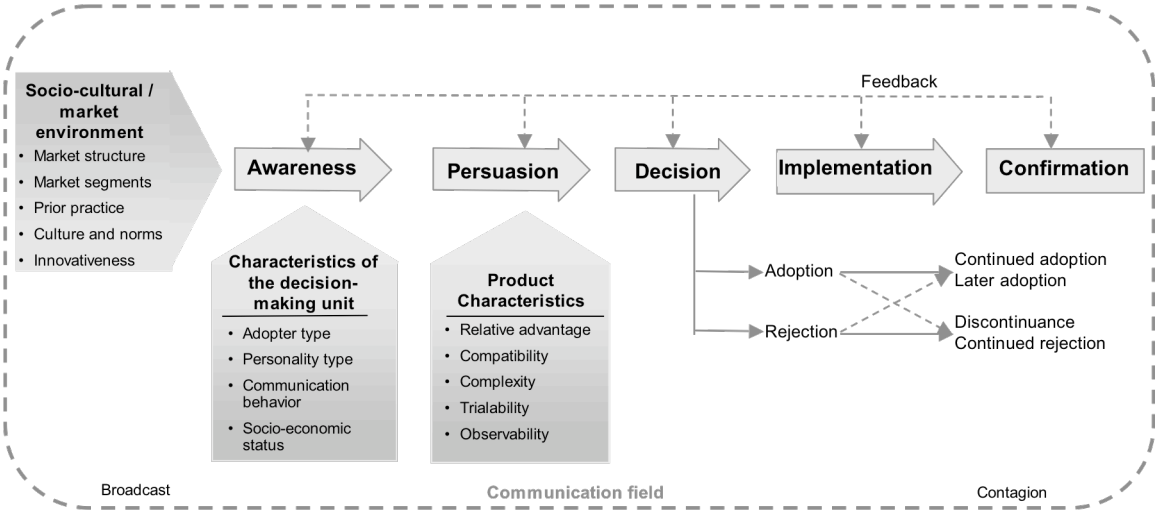
2. Rogers' Diffusion of Innovations

Rogers' ideas help to explain how programs really work to create change.

Diffusion of innovations is a term that came to prominence with the publication of the first edition of Everett M. Rogers' book, *Diffusion of Innovations*, in 1962. The diffusion of innovations is a comprehensive and fairly detailed theory that describes how ideas and technologies find their way into markets and cultures. The theory has evolved over the years, so that Rogers' book is now in its fifth edition (Rogers, 2003). The theory is based on thousands of research studies of interventions from the fields of human services, public health, communications, marketing, sociology, political science, health, and others. It has been used as a basis for the design of literally thousands of intervention programs, although it has not been widely used in the fields of energy efficiency and renewable energy. Diffusion of innovations theory captures much of what is known about social change and how to create it.

2.1 An Overview of Diffusion

Figure 3 is a schematic of the diffusion of innovations theory. It includes five major elements. The *diffusion process* is represented as five stages with the large arrows in the center of the diagram, the adoption/rejection decision at the lower right, and the dashed lines and arrowheads at the top of the diagram representing feedback among the stages. The *sociocultural* part of the model, which deals with the market environment, is represented at the upper left. The dashed gray line surrounding the central diffusion model represents the *communications field*. *Adopter characteristics* are below the awareness stage, and *product characteristics* are below the persuasion stage.



Source: Everett Rogers 1994 as modified by Innovologie, LLC. 2005

Figure 3. Rogers' Diffusion of Innovations Model

2.2 Diffusion Process Model

The basic diffusion model is comprised of five stages that represent how people or firms adopt an idea or technology:

How people react to something new:

- *We find out about it.*
- *We learn more about it.*
- *We decide whether to try it.*
- *We implement our decision.*
- *We see if we like it.*
- *We may or may not try it a second time.*

- Awareness — The individual or firm – the "actor" – becomes aware of a technology or practice. Awareness may come through interactions with others, from personal observation, or from broadcast sources such as newspapers, television or publications.
- Persuasion or information seeking — The actor considers a technology or practice and collects and evaluates information about it. Individuals may be aware of an idea or a technology long before anything is done with the information. The information may flow over the individuals until, at some point, the information may become salient, and the individual or firm begins to consider the technology or practice. Information typically comes with a valence, e.g., a technology or practice is good or bad. Given enough information with the same valence, the decision maker may be predisposed to accept or reject the product or idea. The information search and evaluation process may be simple and quick or calculated and drawn out. The process is informed by the decision maker's values and ideas and the sociocultural context.
- Decision making — The actor makes a decision to adopt or reject a technology or practice. This decision is typically separate from the actual use or implementation of the technology or practice. The decision to adopt a technology or practice may occur months or years before the technology or practice is used or implemented.

- Implementation —The actor implements the decision to adopt the technology or practice by purchasing or using it. The first implementation may be a trial use. For example, the adopter may try buying one compact fluorescent lamp, or auditing one building, to see if the technology or practice is suitable.
- Confirmation — Actors implementing a decision to adopt typically confirm their decision. They may decide to purchase more of the technology, use the practice again or more often, or reject the practice or technology. An example of this process is an industrial firm that has an assessment done, installs some energy efficiency measures, and then takes measurements or has another assessment done to confirm that the measures work. The adoption decision is subject to repeated confirmation attempts, and there is the potential for temporarily or permanently rejecting the innovation.

People may go through the process several times.

The process is assumed to be iterative as illustrated by the dashed lines and arrows above the stages. A fundamental assumption is that individuals or firms continuously re-evaluate their relationship to technologies and practices (innovations). The actor may initially perceive the innovation to be of low salience and reject it but later return, search for more information, and make the decision to adopt. Information may be objective facts about the innovation or may be comprised of perceptions about the innovation derived from interactions with others.

Change can be tracked by using a few simple questions to identify the stage of adoption through which actors are passing.

Because this is a stage theory, the penetration of an innovation can be tracked by examining the stage of adoption for individuals or firms in a population. For example, an actor can be assigned to a stage of diffusion if questions such as the following are asked in a sequence and terminated when a negative response is received.

1. Have you heard of the innovation? (Awareness stage)
2. Have you tried to find information about the innovation or talked with friends or colleagues about it? (Persuasion or information stage)
3. Have you decided to try the innovation? (Decision stage)
4. Have you actually purchased, used, or installed the innovation? (Implementation stage)
5. Have you decided to continue or considered continuing to use the innovation? (Confirmation stage)

One of the values of monitoring the progression of a population segment through the stages of adoption is the ability to determine if a deployment program is synchronized with the stages of adoption for the target population. For example, if awareness is high but program activities are oriented to awareness rather than persuasion or implementation, the program may not be effective.

2.3 The Physical, Social, and Cultural Environment (Market Model)

It is important to understand the structure of the market so that you know where key decisions are made and can influence them.

The model emphasizes the need to analyze and understand the physical, social, and cultural environment in which diffusion of a technology or idea is taking place. Important questions that can be applied in each of the domains are:

- Who are the players?
- Who are the key decision makers?
- How do the players relate to one another?
- What is the context within which the players operate?
- What can be done to facilitate adoption in this context?
- Are there social, structural, or cultural barriers that prevent adoption of the technology or practice?
- What are the interconnections between the players in the different domains?

Technologies and practices sometimes fail because they are poorly conceived. However, they are probably more likely to fail because the technology or idea is inconsistent with the environment in which it is to be used, or because it is poorly marketed. For example, the appliance market is a national market. Manufacturers are more likely to respond to broad based rather than local interests. Thus, local programs targeting appliance efficiency may have limited success in getting manufacturers to increase the efficiency of their appliances. Broad based regional and national efforts may be required to influence key players. Advocates of efficiency have sometimes used the existence of local variations to encourage manufacturers to adopt higher levels of efficiency. The same is true for large production homebuilders and large

developers of commercial properties who operate in multiple jurisdictions (Reed *et. al*., 2004b).

Cultural and social norms, such as local codes or union rules, may impose requirements that make it difficult or impossible to install an innovation or that limit its benefits. The current call for common standards for net metering is a good example of the way in which local norms can slow the diffusion of technologies and practices. Every state has, or doesn't have as the case may be, its own net metering rule.

Cultural artifacts – for example, existing equipment – may inhibit or slow the adoption of innovations. The early problems of installing compact fluorescent lamps in floor and table lamps and ceiling fixtures are good examples. The availability of dimmable compact fluorescents with a high color temperature would likely result in a much more rapid diffusion of compact fluorescent bulbs. Understanding and facilitating the adoption of an innovation requires analysis and understanding of the social, political, and cultural environments (market) in which the adoption is to occur.

Cultural artifacts, such as rules, regulations, and norms, can prevent or facilitate change.

2.4 Communications Model

The communications model suggests that communication takes one of two forms. *Broadcast* is a one-to-many process – i.e., there is a single source with many recipients of the message. Advertising, magazine articles, mass mailings, bill inserts, and websites are examples of broadcast methods.

Contagion is the movement of information from person to person. Contagion occurs through networks. People pass information to each other within their own network. Those who are more marginal to a network and who are members of other networks tend to pass information from one network to another.

Because in our culture we are surrounded by mass media, we are somewhat conditioned to think of broadcast processes as being the more efficacious of the two processes. Companies spend billions on advertising. Advertising

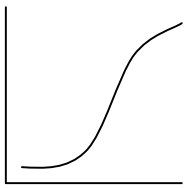
does work. A strong advertising campaign for a mature product can sway usage by a percentage point or perhaps even a few percentage points. If you are selling sugar water, that is a lot of money.

We are bombarded with a lot of media advertisements, but personal communication – what we say and what we observe about each other – is usually the most effective way to spread the word.

However, contagion is really the more important of the two processes in terms of reaching the largest number of people. There are a large number of books and articles that point to the centrality of networks and "creating buzz" as a product marketing model (Rosen, 2000).

It is possible to model and project market penetration using various mathematical formulations of the "S-curve".

Many people's understanding of the diffusion of innovations is limited to the "S-curve" that reflects the communication part of the model. Bass and others have shown that the diffusion curve, as the S-curve has come to be called, can be mathematically modeled by choosing parameters to represent broadcast and contagion processes (Bass, 1967). Models can be developed to incorporate the effects of incentives and other market activities and features. Also, the trajectory of the curve can be determined from a few points, making it possible to estimate future penetration levels. Parameters from other technologies can be used to estimate the trajectories of technologies or practices with no history (Mahajan, 1985).



We tend to forget the technologies and practices disappear. Sometimes program implementers want to rid the market of a technology, for example, inefficient refrigerators or cars with low gas mileage.

Less attention has been given to the backside of the S-curve, i.e., the leveling off (or decline) of the adoption curve. At some point technologies are overtaken by similar technologies with more desirable features or other innovations. The decline may be slow at first and accelerate as the new technology becomes more desirable and accessible. An example of technology evolution and displacement would be black and white televisions that were displaced by color televisions, which in turn are being displaced by large screen televisions and high density television (HDTV). Consider phonograph cylinders, vinyl records, compact diskettes, and MP3 players. These are successive generations of technologies. Sometimes programs are implemented to accelerate the removal of products or generations of products, for example, utility programs to remove older less efficient refrigerators.

There is an entire literature on technological change and how technologies are created, rise, and fall. One of the key issues in that discussion is how to classify technologies as "evolutionary" or "disruptive" (Mollas-Gallart, 2006). The advent of television might be considered a disruptive event because it significantly reshaped communication. The changes from black and white to color to big screen and HDTV are largely evolutionary changes. They did not significantly alter how communication occurs. The movement from phonograph cylinders, to records, to compact disks is probably also evolutionary. The advent of hand-held MP3 player and Internet downloads is likely disruptive because it is reshaping the commercial arena by reducing the importance of music stores, potentially increasing access to independent musicians and video makers, and moving the delivery of video content from a scheduled activity to a demand-based activity. For an evaluator, it is important to know if the technology or service is evolutionary or disruptive. If the technology is disruptive, the market transformation impacts of such a technology or service may be far reaching and quite different.

A lot of the technological change is evolutionary, consisting of incremental improvements to existing technologies.

From time-to-time a disruptive technology appears that is different than what came before and causes significant rearrangement of social and cultural patterns.

Many of the early diffusion of innovations studies revolved around understanding, developing, and using S-curves. More recently, there have been breakthroughs in our understanding of networks. Network analysis is potentially a powerful tool for understanding and accelerating the pace of diffusion, such as contagion (Barabasi, 2002; Backstrom et. al., 2006; Borgatti, Everett, and Freeman, 1999; Granovetter, 1973; Granovetter, 1974; Johnson S., 2001; Kempe et. al., 2003; Kempe et. al., 2005; Klovdahl, 1985; Leskovec et. al., 2006; Milgram, 1967; PAJEK; Reed and Oh, 2003a; Reed and Oh, 2003b; Rheingold, 2002; Scott, 1991; Scott, 2000; Taylor, 2001; Watts, 1998; Watts, 2003; Whyte, 1954). For the evaluator, network analysis may become an indispensable tool for program evaluation. Currently it is useful for identifying who is in networks and how they are related. Its importance will increase as we learn more about how to analyze and interpret the content of linkages.

A potentially important tool for impact evaluation is network analysis. It can help us understand how technologies or practices spread.

2.5 Characteristics of Adopters (Model of Adopter Types)

People have different tolerances and willingness to try the new. We use terms like innovators, early adopters, early majority, late majority, and laggards to categorize such people.

Another aspect of the diffusion of innovations that is widely known and frequently cited is the adopter types. Innovativeness is a personality or organizational characteristic that influences actors' receptivity to new products and ideas. Figure 4 shows Rogers' (2003) classifications of actors in terms of innovativeness and the approximate percentage of the population that he estimated for each category.

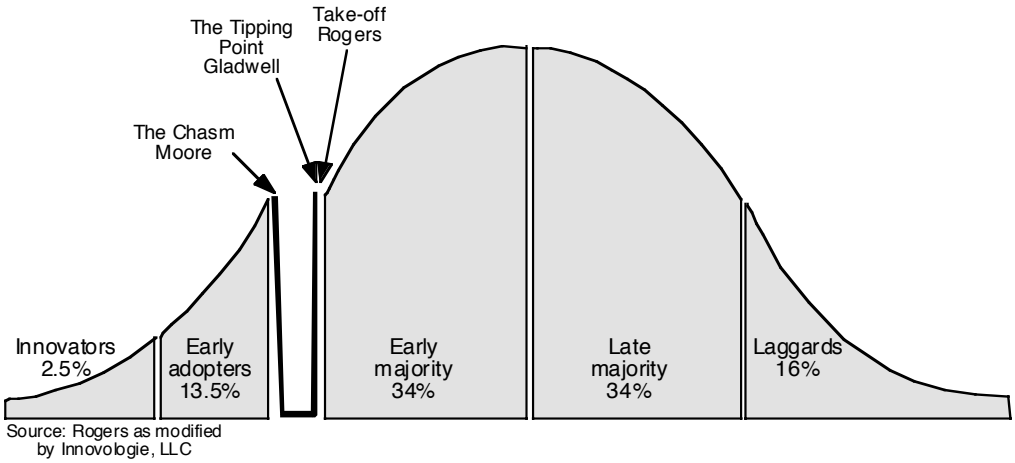


Figure 4. Categories of Adopters and Their Approximate Distribution in the Population

- *Innovators* are constantly looking for new innovations and are the first to try them. They purchase technology or try new ideas out of pure interest in the technology or the idea. They are willing to tolerate flawed and "half-baked" technologies. They obtain information about innovations through mass communications channels. They are not afraid to abandon products that they perceive not to have long term value. Innovators often play a gatekeeper role by controlling and/or influencing the flow of ideas into the social system. They are important because their use of the product can lead to continued production and refinement.
- *Early adopters* are typically respected members of the community. Like the innovators, they get their information from mass communications channels. They appreciate the potential benefits of technology, and they utilize the technology when its benefits match their own needs and desires. They are more integrated into the social system than innovators. They are a bit more judicious in the innovations that they adopt. Their choices are important as examples for others in the

social system. People seek out early adopters for their recommendations.

- The *early majority* is the core of the population that must be attracted in order for a product or idea to make it fuller into in the market, that is, for the market to be "transformed." They are inclined to wait until an innovation is proven and then adopt it. They tend to reference each other rather than innovators and the early adopters. Interpersonal communication (networking) is much more important to their decision to adopt than broadcast channels. Their acceptance of an innovation is critical to the continued adoption of a product or idea. Their adoption signals that market transformation has truly taken hold.[1]

  The "chasm" between early adopters and the early majority is a crucial crossover point. Rogers (2003) labels this crossover point the "point of take-off." Moore (1991) refers to this as "crossing the chasm." Gladwell (2002) refers to it as the "tipping" point. What innovators and early adopters use or find interesting does not easily or necessarily translate to adoption by the early majority. Getting the early majority to adopt an idea requires that interveners get respected parties from among the early majority to adopt the idea. Other individuals and firms will then reference these adopters and take up the technology for themselves. If a change agent is targeting grocery chains, it is not just a matter of getting just any grocery chain to adopt a technology but of getting a chain that other chains will emulate. Thus, it is important to carefully select targets for change. An innovation or idea that works with innovators and early adopters does not necessarily guarantee success with the early majority. This is why products sometimes seem to take-off and then suddenly lose traction in the marketplace.

Innovators and early adopters are very different than the early and late majorities and laggards. Broadcast strategies work with innovators and early adopters, but contagion is more likely to be successful with the early and late majority.

- The *late majority* also represents about 34 percent of the population. They are not comfortable with technology and they tend to be risk averse. They will wait until a product or idea has become the standard before they adopt. They may not have the resources to try innovations. Pressure from peers leads these firms or individuals to adopt new ideas.

- The *laggards* are about 16 percent of the population. They tend to continue what they have done in the past. They may be suspicious and even hostile to innovation. They are generally the last to adopt and may not adopt until it is no longer possible to continue to use old products or ideas.

[1] In this context, "market transformation" means that the market is changing, and these changes are expected to last, as more and more individuals and firms implement a technology or practice.

2.6 The Product Model or the Characteristics of the Innovation (Product Characteristics Model)

The specific characteristics of a technology or practice generally determine whether and how it is adopted.

Some technologies and practices have characteristics that facilitate their adoption, while others have characteristics that may impede their adoption. The five characteristics of technologies or innovations that determine whether a product will have an easy or difficult time are:

- Relative advantage
- Compatibility
- Complexity
- Trialability
- Observability

A lot of people think adoption is about price, but it is really about relative advantage of which price is just a small piece.

An innovation is more likely to be adopted if it confers ***relative advantage*** over other products or ideas. For example, an innovation is more likely to be adopted if it is less expensive, produces more at lower cost, or has other features such as style, ease of use, or status-conferring properties that make it attractive vis-à-vis other products or ideas. Although much is made of price, price is only one factor and perhaps not even the most important factor in the decision to adopt technologies or practices. Consumers are rational, but consumers are rational across a number of characteristics not limited to price.

A technology's compatibility with systems, with markets, and with culture is often overlooked.

The possibility that an innovation will be adopted is increased if the innovation is ***compatible*** with the social, cultural, and physical environment in which it is to be used. The phrase "plug and play" is a manifestation of this in contemporary culture. If one has to modify the existing environment – e.g., change building codes – or purchase new equipment to utilize the innovation, then the likelihood of adoption decreases.

The current iPod craze is facilitated by the fact that Apple created a set of software that makes it easy to transfer music from CDs and allows the purchase of music and video online. Without these innovations, the iPod would likely be just another gadget in the drawer. While video downloads already exist, their use is likely to accelerate as high-speed Internet connections and hardware that allows their easy use anywhere in a household or office become

more ubiquitous. The point is that technologies and practices do not stand alone but exist within ecosystems that support them. Compatibility (complementarity) is a key issue.

Complex innovations, or innovations that are perceived to be complex, are less likely to succeed than are simple innovations. The adoption of many electronic and computer products is slowed by their ***complexity***, real or perceived.

Innovations that can be easily tried – have high ***trialability*** – have increased chances of adoption. We constantly receive free samples in the mail. We are bombarded by advertisements urging us to get a trial pack of a pharmaceutical from our doctor. We are stopped in the aisles of stores and offered a taste of some food or a spritz of cologne. People who try an innovation are more likely to adopt the innovation. We know that the large national builders and retailers often use a new product in a single building and then in a subdivision to see what potential problems arise and how it works before adopting the product for all of their projects. The same is true of national supermarket chains (Reed, Johson, *et. al.*, forthcoming 2007; Johnson, 2007b.

A lot of firms like to pilot a technology before they attempt to install it everywhere. High trialability can improve the chances of adoption.

Finally, there is ***observability***. An innovation is more likely to be adopted or not adopted if the benefits or disbenefits of using the innovation can be physically observed. The effects of new lighting or how well a set of controls adjusts lighting is observable, especially if lights go out while one is sitting quietly in one's office, as sometimes happened with early lighting controls. Reduced energy use is much harder to observe. Developing ways to make the effects of an innovation observable can accelerate the adoption of an innovation.

There are now several companies that collect and organize utility bills for large businesses with multiple outlets. Some of these companies also provide a service that compares the monthly energy use of a building to other similar types of buildings (a building baseline). This comparison makes the energy use of the building observable. In turn, it can lead to increased attention to

energy efficiency. The more observable something is, the more it is likely to be adopted.

2.7 Why Diffusion of Innovations

You might ask, why diffusion of innovations and not some other approach? The goal of this framework is to provide a broad systems theory that can provide the underpinnings for a more unified approach to evaluations across programs. In addition, the framework increases the potential for programs to undertake collaborative evaluations and to make it easier to aggregate benefits. In the absence of a centralized framework to guide the development of measures, evaluations largely become a series of case studies, which while useful, are difficult to knit together into a more robust understanding of overall portfolio effects. The idea is to be as broadly inclusive and as systematic as possible.

Megdal has argued that the energy evaluation community should take advantage of other behavioral and social theories (Medgal, 2005).[2] We agree. The problem isn't one of choosing among theories. Rather the problem is one of integrating theories to develop richer understandings. Thus, we have chosen diffusion of innovations because of its breadth and because of the need to provide a consistent framework across programs. Other theories can

[2] In this regard, a number of people have asked how Fishbein and Aizen's (1975) theory of behavioral intentions relates to the framework. Does this framework embrace it or displace it? Behavioral intentions evolved out of frustration with large numbers of experiments in the 1940s, 50s, and 60s that revealed no correlation, or only a small correlation, between attitudes and behavior while finding strong correlations between antecedent and subsequent behaviors. The theory states that beliefs and attitudes predict behavioral intentions that predict behaviors and that there may also be a direct link between attitudes and behavior. Intentions were offered as the intermediate step between attitudes and behavior to get around the lack of correlation between them. More recently, Aizen (1991) described a theory of planned behavior that removes beliefs but includes attitudes and adds subjective norms and perceived behavioral control.

The persuasion, decision, and implementation stages that are part of the diffusion of innovations essentially address the same set of issues but escape the trap of having to deal with beliefs and attitudes. Attitudes, subjective norms and personal behavioral control are analogous to the persuasion stage which is fed by product characteristics, for example, "relative advantage" that is a form of subjective norm. "Perceived behavioral control" could be an adopter characteristic and the product characteristics of observability and trialability readily feed perceived behavioral control.

The decision stage is analogous to behavioral intentions. In fact a number of definitions of behavioral intentions have the word "decision" in them. A recent paper from the National Cancer Institute (2007) states, "Decisions to act or not act are the result of an assessment of the likelihood of specific outcomes associated with the act along with the subjective value assigned to those outcomes. When the assessment produces a positive evaluation, a decision is made (usually) to act." Finally, implementation is the behavior stage. Thus, much of the Aizen's theory is incorporated in quite specific ways into diffusion of innovations.

usefully be integrated to enhance understanding of the diffusion of innovations.

3. Framework for Impact Evaluation

In this chapter, we present a framework for completing an impact evaluation. In later chapters, we focus in more detail on various aspects of the framework.

The framework has seven modules or steps:

1. Identify the scope, objectives, and priorities.
2. Select the type(s) of evaluation to be completed.
3. Select aspects of deployment changes to be evaluated.
4. Identify researchable questions and metrics.
5. Design the study and select the methods.
6. Conduct the evaluation.
7. Report and use results and data.

Figure 5 provides a more detailed view of the framework, especially additional elements associated with each of the modules that are discussed below.

3.1 Module 1: Identify the Scope, Objectives, and Priorities of the Impact Evaluation

There are never enough resources to do all of the evaluation that a program manager might like to do. This puts a premium on identifying the scope, objectives, and priorities of an impact evaluation. With a clearly delineated scope, objectives, and set of priorities, a manager can use the available resources to produce useful evaluations.

Identifying the scope, objectives, and priorities of an impact evaluation for a deployment program can be broken down into five steps.

1. Describing the program by updating or creating a new logic model.
2. Setting the type of evaluation to be completed.

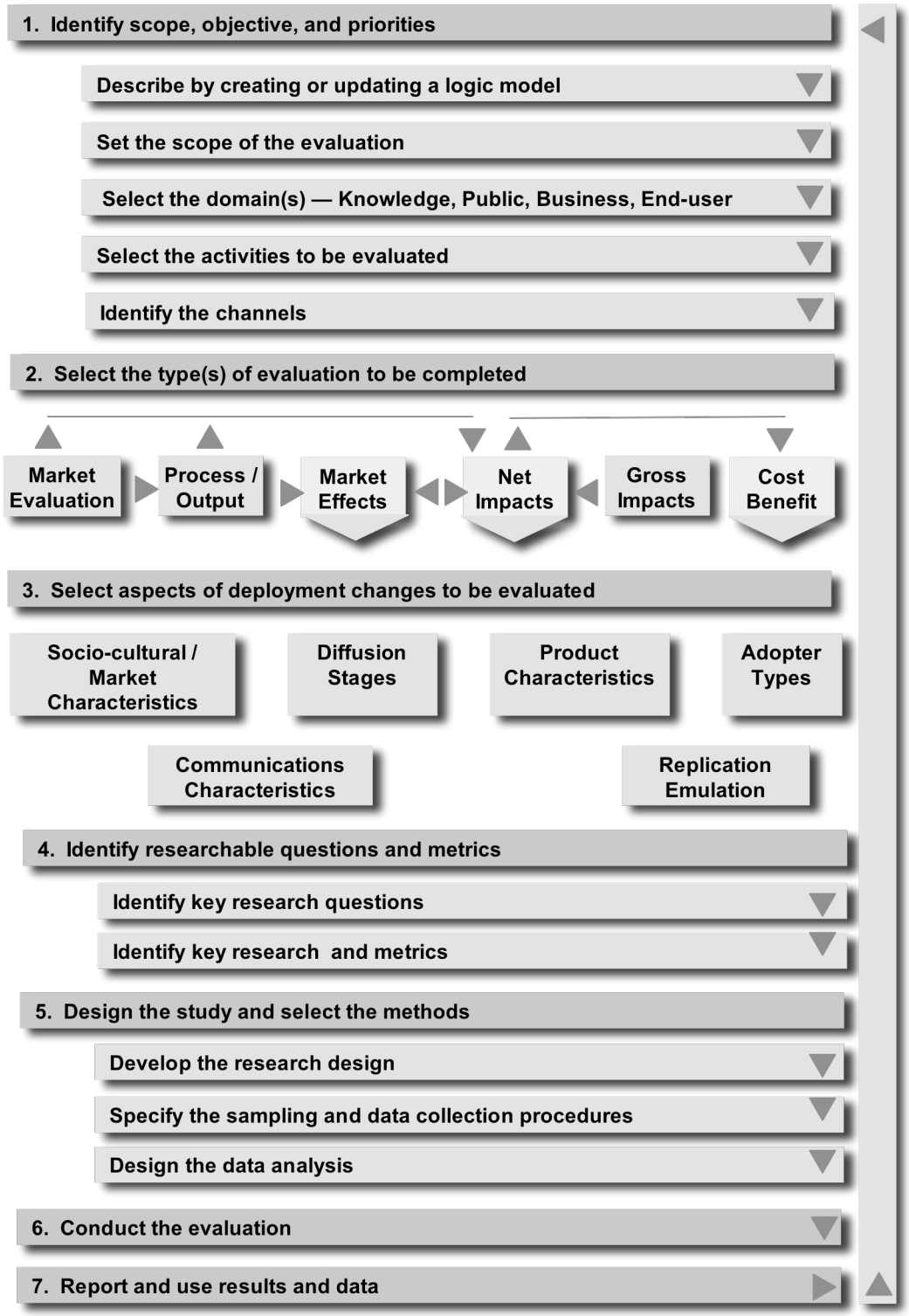


Figure 5. Overview of the Impact Evaluation Framework

3. Selecting the domain and targets within the domain that are to be the subject of the evaluation.
4. Select the activities to be evaluated.
5. Identify the delivery channels to be evaluated.

3.1.1 Describe the program by updating or creating a new logic model

We cannot emphasize too strongly how important it is to have a good description of a program before attempting to design an impact evaluation. The description must include a clear statement of the goals, objectives, and strategies associated with a program. The goals should be clearly stated in a "what is expected to be accomplished by when format." A good description of the program tells how the "what" is to be reached in the expected time frame and makes it easier to identify the aspects of a program that need to be evaluated. A good description also helps to assure that you identify all of the factors that may need to be taken into account in the evaluation.

State goals, objectives, and strategies clearly. A good format for a goal statement is "to specify what will happen by when."

Periodically or at the end of an evaluation, the evaluator and the program manager should revisit the goals. The purpose is to assess whether the content of the evaluation goals is appropriate in light of the program goals and whether the program goals are realistic or whether they need to be re-evaluated.

One of the best ways of generating a good program description is to update or create a program logic model. A logic model is usually a two dimensional figure with a logical sequence of program activities in one dimension and the resources, activities, outputs, and short and long term outcomes associated with each of the activities arrayed in the other. Some programs have logic models while others may not. Even if a program has an existing logic model, it is important to re-examine it before starting an evaluation and ask how the program has changed since the logic model was last updated. Program managers are constantly making changes to their programs for a variety of reasons. For example, they come up with a new approach, discover that they cannot do something that they thought they could, improve some-

Update or develop a program logic model to describe your program.

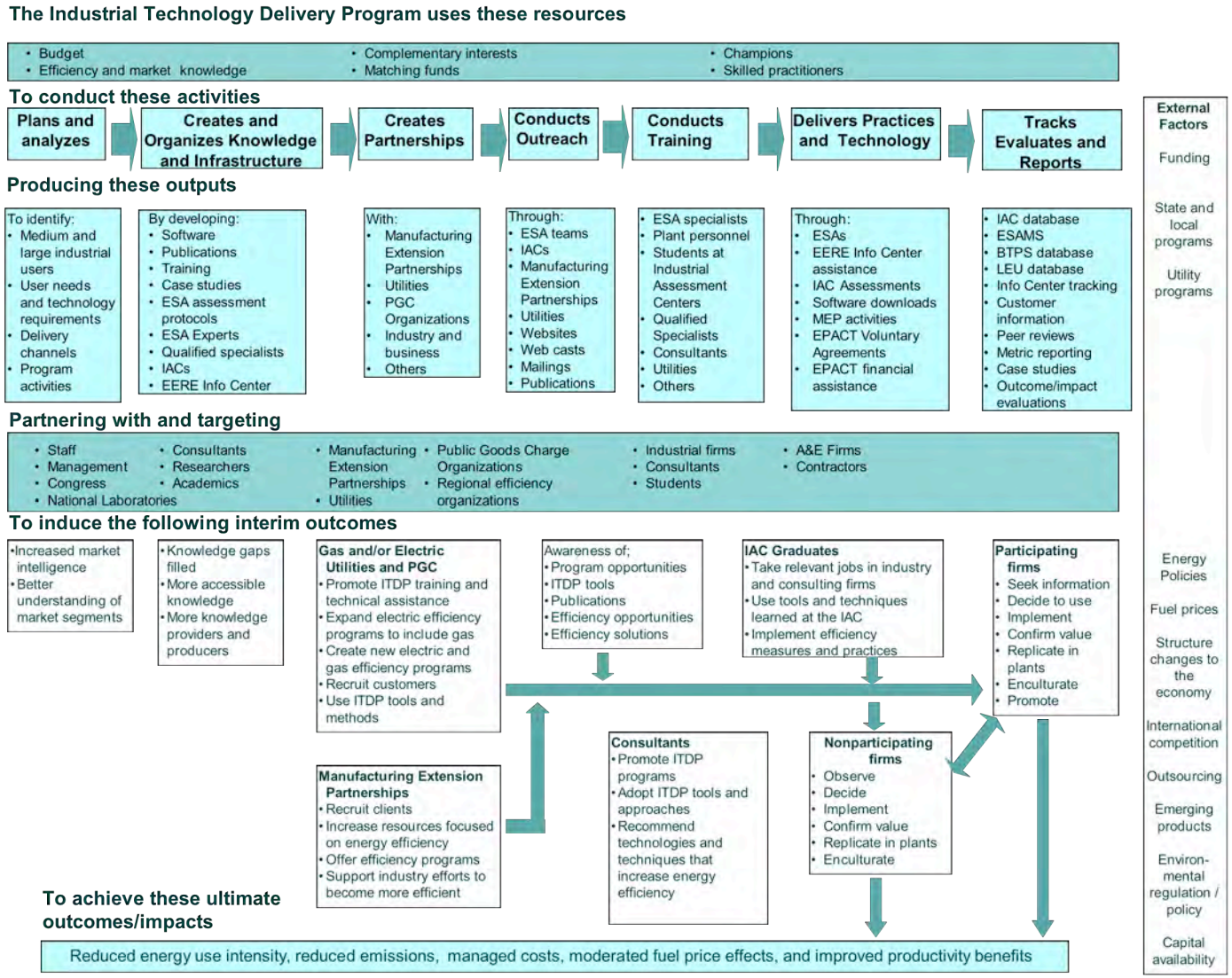
thing that has not worked, or discover forces at work in the external environment that need to be leveraged or taken into account.

A logic model helps simplify the understanding of complex programs, shows the linkages among elements, and highlights what you expect of target audiences and partners.

Figure 6 is one version of a high-level logic model for EERE's Industrial Technology Delivery Program (ITDP). A logic model like this is very helpful because it provides evaluators with a lot of program information. One striking fact about the ITDP is the complexity and the many inter-connections among its various components. For example, the program has an activity where it creates software and a variety of other activities that result in the software being used, for example, basic software downloads or training programs.

More importantly, the logic model helps evaluators quickly identify the large range of outcomes that can result from program outputs. Examples include a student learning to use the software while doing an industrial audit, a graduate of the Industrial Assessment Center (IAC) program using the software for her employer, an industry consultant downloading the software to see what might be improved in his own spreadsheets for analyzing steam systems, a facility engineer looking for ways to control costs, and so forth. Understanding these pathways can help the program manager to more fully identify their potential impacts and to identify evaluation priorities.

Logic models are sometimes criticized because feedback loops are not typically shown. People who use and develop logic models understand that the interconnections and feedback loops are implicit. For example, it is understood that feedback from architectural and engineering partners to software developers in the knowledge domain might be an outcome of training, or that use of the software could lead to changes in the software. Part of the job of an evaluator, and an important use of logic models, is to pinpoint areas where feedback needs to occur and ask if the organizational connections are in place to facilitate the feedback.



Source: ITDP and Innovologie LLC. As of March 2007.

Figure 6. Industrial Technology Delivery Program Logic Model

Logic models are often used to describe a view of how a program currently works. That is a useful first step. Some suggest that this is a major limitation of logic models because it may discourage program managers and evaluators from thinking beyond the obvious.

Logic models are much more useful when they are forward looking, anticipating what actors, what actions, and what connections between actors are required to accomplish program goals. Logic models can be used to analyze what needs to occur as well as what is occurring. Indeed this is one of the main themes in this document. It is easier to think about what the other connections may be if you have a story line or theory to help guide your thinking.

In addition, useful logic models identify external factors that impact the efficacy of programs. In the ITDP model, outsourcing is listed as an external factor. There has been a general trend for industrial firms to use more outside engineering staff than in the past. Managers in industrial plants report that this leaves fewer internal personnel to identify, promote, manage, and implement energy efficiency projects. It may also make it more difficult to maintain gains in energy efficiency. Identifying external factors is quite different than describing the current state of the program and should not be overlooked.

3.1.2 Set the scope of the evaluation

Program managers need to ask, "For who are we doing the evaluation and how are we going to use the results?" The results of impact evaluations can have many uses. For example, they can be used to:

Tailor the scope of the impact evaluation to the needs of your program and stakeholders.

1. Defend or justify expansion of a program by showing that the program delivers savings, reduces emissions, or has other impacts.
2. Decide where to allocate dollars to produce the most savings.
3. Decide how to increase the efficacy of a program component.
4. Provide estimates of performance needed for The Government Performance and Results Act (GPRA), OMB PART, etc.

If the purpose is to defend or expand a program, a place to start is to ask what activities produce the most savings. If a program has already conducted some evaluations of savings, it is appropriate to ask what's next to be evaluated. Or, if the amount of money spent on training is being questioned, it is important to examine the energy savings from training and compare that to the cost of producing the savings.

If the goal is to decide where to allocate dollars, it is important to do three things. First, decide what activities are in competition so that choices can be made among them. Second, for activities that are not in competition, ask what elements are essential to the success of the program. And third, prioritize the activities based on savings.

If the goal is to increase the efficacy of a program component such as a training program or an energy audit, it might be useful to design an experiment to try different approaches and to evaluate the differences in savings.

The bottom line is that one has to answer two questions:

- Why and for whom am I doing this?
- How am I going to use the results?

3.1.3 Select the domain and the target actors

Narrow the impact evaluation to focus on the important domain(s).

In the introduction, we mentioned domains and target actors. Table 1 lists some typical actors in each of the four domains. The list is not exhaustive but it is a useful aid for thinking about who to target as part of the impact evaluation. Also, the actors can appear in more than one column, meaning that they can be a target, a partner, or both.

Many deployment programs focus on multiple domains — knowledge creation and packaging, public policy or nongovernmental organizations, industry/business, end-users — and multiple participants within the domains. For example, the commercial buildings program may want to target retailers as end-users: for example, the program may want to influence the retailer to

build more efficient stores. The Energy Star program may target the retailer as a business partner or ally to promote the sale of efficient products.

Table 1. Actors in Four Domains

Knowledge Firms	Policy Makers and Public Entities	Manufacturers / Businesses	Energy End-Users
Laboratories Universities Research labs Consultants Media specialists Web developers Modelers	Regulatory commissions State energy offices State code officials Legislatures Public goods charge programs Regional efficiency organizations Nongovernmental organizations Community organizations Agricultural extension service Manufacturing Extension Service	Materials suppliers Product manufacturers Distributors/wholesalers Utilities Retailers/builders Architects Engineering consultants Specifiers Service providers i.e., auditors, HERS* raters, etc. Installers/contractors Maintenance contractors	Households Commercial building owners Tenants in commercial lease building buildings The buildings of retailers The facilities of industrial firms Vehicle fleet managers

*HERS is the Home Energy Rating System

From an evaluation perspective, the logic should include an examination of how the program relates to each domain. Having done this, the evaluator can then identify the domains and the roles and relationships on which the evaluation is to focus. If a program is partnering to leverage the activity of a business, then the evaluation needs to focus on how the program influences its partners. For example, the ITDP trains consultant specialists. Thus, the ITDP may want to focus the evaluation on what the consultant does with the training, whether the consultant uses the training, how the consultant uses the training, and what actions result from the use of that training. Does the consultant use the training on all of the projects, some, or none? What are the energy savings when the consultant uses the training? Is the consultant replicating the use of the training when dealing with other firms? If so, what is the effect?

3.1.4 Identify the activities to be examined during the evaluation

The ITDP logic model identifies seven main activities with a number of outputs giving indication of a large number of sub-activities. For large programs like those found within EERE, there are never sufficient resources to evaluate everything at once

The program manager and evaluator need to identify the activities that are to be evaluated. In general, it makes sense to evaluate activities where the least is known, the stakes are high in terms of the overall goals of the program, stakeholder interest is high, and resource allocation is potentially an issue. If not enough resources are available to complete all the evaluation activities in a single year, then a multi-year strategy can be used.

Think multi-year to optimize resources!

The key is to decide which of the main activities and which of the sub-activities should be evaluated and the order in which they should be evaluated. If the focus is on the ITDP activity "delivers technologies and practices," there are many sub-activities. So, assuming the evaluation is limited to a focus on training, there are still numerous possibilities. Some examples are:

Think activity!

- Evaluate all technical assistance and training programs.
- Evaluate just training programs.
- Evaluate the training system including content development (under the infrastructure activity) and training delivery.
- Evaluate just one training program.

Evaluating the impacts of all training and technical assistance programs would allow the program manager to see the relative savings among all these sub-program elements. Likewise, evaluating all training programs would allow the examination of the savings across different training programs. If the training programs to be evaluated were well chosen, evaluating just two or three training programs might allow the manager to get a feel for what a small or large impact might be or how much replication, emulation, spillover, and sustained behavior are occurring.

Consider joining with other programs to evaluate common activities such as training and websites.

Another possibility to keep in mind is cross-program evaluation. For example, most programs within EERE offer training. Thus, a large, joint training program evaluation might be a possibility. In another example, many programs have websites. Several programs might want to band together to examine the effects of their websites. Each program might be able to contribute a smaller amount of resources that would result in a larger, more productive evaluation than a stand-alone evaluation of a website. The diversity of the websites might make it possible to evaluate more elements increasing the utility beyond what could be learned from evaluating just one website or the websites that belong to one program.

Again, the issue boils down to prioritizing activities in terms of where you need more information, where impacts may be large or potentially large, where resource allocation may be an issue, where stakeholder interest is high, and trading off these factors against evaluation resources and the potential for doing a sequence of evaluations over multiple years.

3.1.5 Identifying the channels to be evaluated

Deployment programs typically use numerous delivery channels to implement programs. For example, awareness may be increased and information delivered through the Web, publications, CDs, brochures, call centers, video conferences, conferences, workshops, training, and other media. Technologies and practices can be delivered directly and indirectly through a variety of channels. Programs may provide direct training on a tool, technical assistance such as architectural design assistance, audits or analysis, financing, incentives, direct installation of measures, and a host of other mechanisms.

A key reason for assessing delivery channels is to understand their effectiveness and/or cost effectiveness. For example, a program may wish to determine whether developing and distributing case studies is an effective way of stimulating target audiences to respond. The content of the case study may influence its effectiveness that needs to be taken into account. The same channel may be differentially effective depending on the content it carries.

Investigating the cost effectiveness of the delivery channel is particularly important because different channels may have result in different levels of response and impacts per participant. A website might have a high number of visitors but cause few actions or actions of low consequence.

The interactions among channels are critical as well. The aforementioned website may might have a low benefit to cost ratio and but may be a critical channel by which participants become aware and get information about a technical assistance program that has high per participant impacts.

Partners typically play an important role in delivering information or measures. Partners may include contractors, colleges and universities, extension services, other federal agencies, state and local governments, nongovernmental organizations, standards setting bodies, utilities, manufacturers, distributors, retailers, installers, and others. There are whole bodies of literature on partnership and community programs as delivery mechanisms. A great deal of this literature is found in the public health field and in the agriculture extension literatures. Key lessons from those literatures are that partners should be chosen carefully for their capabilities including organization depth and relevant experience, effectiveness in working with intended target groups, and positioning in the community.

It is important to examine partner relationships to understand what partners are contributing. Building and maintaining partner relationships can be very expensive and, while producing great satisfaction for the partners, there are documented instances in utility settings where partner relationships have not proven productive.

It is unlikely that a deployment program would evaluate all of its delivery channels at one time. Thus, it is important to establish some priorities. A process for might be as follows:

1. Identify the delivery channels being used by the program.
2. Determine the purpose of the delivery channel.
3. Assess the outputs and outcomes expected of the channel.

4. Identify channels with large expected outcomes and channels that are necessary for other channels to function.
5. Identify channels where there is substantial uncertainty about outputs and outcomes.
6. Examine the expenditures on the channels.

The various factors may be traded off to determine which channels to evaluate. Priority might be given to those channels where expected outcomes are high, the cost is great, and there is substantial uncertainty about what the channel actually delivers in terms of outcomes. In later years, the evaluation might examine other delivery channels.

3.2 Module 2: Select the Type of Evaluation to Be Completed

In this framework, we are primarily concerned with impact evaluation. There are other types of evaluation: market, process, outcome, and cost-benefit evaluation (See below).

In impact evaluation, we are most interested in net impacts as opposed to gross impacts. Gross impacts are the total impacts from which non-program effects have not yet been subtracted. For example, if the program is promoting compact fluorescent lamps (CFLs), the savings from installing all CFLs that are sold could be estimated. The result would be a substantial over estimate of the program savings, because some people might purchase CFLs without the program, and some may purchase a CFL and then not install it or decide to remove it once installed (Reed, 2003c). If the person uninstalls it, it is likely a case where the person "confirmed" they didn't like it.

Think beyond gross impacts!

Gross impacts can be misleading and are often called into question by those who doubt program success. It is strongly recommended that programs not conduct evaluations that only focus on gross savings.

Program Evaluation Types

Process or Implementation Evaluations

Examine the efficiency and effectiveness of program implementation processes. The results of the evaluation help program managers decide how to improve program operations.

Outcome/Market Effect Evaluations

Estimate the success of outputs in achieving objectives in a specific timeframe. Findings show how well the program is achieving its intended outcomes or creating specific market effects such as convincing trades persons to adopt a specific technology or practice. This helps program managers decide on whether to continue the program and at what level of effort.

Impact Evaluations

Take outcome evaluations one step further by estimating the proportion of the outcomes that are attributable to the program rather than to other influences. As with outcome evaluations, these findings help program managers decide whether to continue the program and at what level of effort, but these decisions can carry greater weight because they are based on findings of causality.

Cost-Benefit Evaluations

Compare program benefits and costs. Cost-benefit evaluation shows the relationship of benefits achieved by program activities to the costs incurred to achieve those benefits. The findings allow program managers to judge the cost-effectiveness of entire programs and of component activities and help them decide whether to retain, revise, or eliminate program elements.

Needs/Market Assessment Evaluations

Assess market baselines, customer needs, target markets, and how best to address these issues by the program in question. Findings help managers decide who constitutes the program's key markets and clients and how to best serve the intended customers. When performed at the beginning of a program, needs/market assessment evaluations also establish baselines against which to compare future progress.

EERE General Program Evaluation Guide (modified) 2006

To assess impacts, it is often necessary to do more than one type of evaluation or at least to complete elements of more than one type of evaluation. Attributing impacts to a particular program may require some elements of market assessment, market analysis, process, or outcome evaluation.

Estimating net impacts is likely to include elements of other types of evaluation, especially process and market effects.

Market baseline analysis may used to determine the channels and the quantity of products flowing through the channels before and after intervention or periodically during an intervention. For example, we may have to determine how many CFLs were sold, how many were sold by what vendors, how many vendors became participants in the program, and how CFLs were sold in response to a program. Sales of CFLs is a market effect from which

energy savings can be determined. Typically in the case of the deployment programs, good impact designs require outcome or market effects evaluation.

A benefit-cost evaluation is the next step beyond impact evaluation in which the benefits of the program are compared to the cost of producing the benefit. A number of states, for example, Vermont, Missouri, California, and Wisconsin have completed benefit-cost evaluations of their weatherization programs. In the Vermont studies, the changes in consumption before and after the homes were weatherized were estimated. The energy savings were then projected over the lifetime of the measures and discounted energy costs were used to project the total weatherization benefit (Table 3). The result was that the benefits of weatherizing the homes exceeded the cost of doing the weatherization. Some studies have examined the benefits of specific measures installed as part of a weatherization program. The goal of these studies is to assess the cost effectiveness of the individual measures.

Table 2. Outcomes from an Evaluation of the Vermont Weatherization Program

Outcome category	Value
Average per unit cost of all activities	$3,451
Net present value of the savings over the 20 year life of the measures	$5,538
Benefit cost ratio	$1.60

Source: Hall and Reed, 2001.

There are two reasons for mentioning benefit-cost studies. The first is to point out that a major component of such studies is an impact analysis. The second is to observe that benefit-cost studies may require other types of supporting evaluations. Generally, people assume that cost data for energy programs are readily available. In fact, cost data are often some of the most difficult data to obtain. The Vermont study was an exception to this because the Vermont Office of Eco-nomic Opportunity had been meticulous in having its contractors keep track of costs at the measure level.

In many organizations, costs are not tracked by program. For instance, in many utilities and governmental organizations, a program manager may be assigned to a program for some percentage of the time, but the actual allocation of hours is not tracked. Further, a project manager may make use of services such as marketing that are not charged back to the project but charged to some overall marketing budget. And lastly, administrative costs for telephone, automobiles, and other services may not be easily tracked. This is less of a problem than it used to be as many large organizations now organize themselves around cost centers and services. A process or implementation analysis may be necessary to identify the sources of all of the costs. The analysis may also have the advantage of identifying activities with costs for which benefits can then be assessed.

You may want to assess the availability of cost data and what it takes to obtain it before embarking on a benefit-cost analysis.

There are four points to take away from this discussion on selecting the type of evaluation.

- The focus should be on net effects and not gross effects.
- Some market effects and even process evaluations are often required (even in resource acquisition programs) to sort out issues of attribution. Some market analysis may be required as well, especially if a baseline is needed.
- Net impact analysis is a key component of benefit-cost analysis.
- Benefit-cost analysis may require some process analysis components to establish cost.

3.3 Module 3: Select the Deployment Changes to Be Evaluated

In the previous chapter, we introduced the diffusion of innovations model. The diffusion model identifies different types of changes that may be stimulated by a program. In developing an outcome or impact evaluation (including market effects), it is useful to consider the various types of changes to ensure that the evaluation will capture everything that is needed.

3.3.1 Sociocultural market characteristics

Deployment programs often spur changes in the habits of end-users and people in industries/businesses, the policy arena and public entities, and

within the knowledge realm. Sociocultural characteristics can reduce or increase the effectiveness of a program by creating constraints or opportunities that make it difficult to adopt a practice or technology or amplify a program's effectiveness. Examples of constraints may be:

- Lack of workers with the right knowledge or interest
- Local codes
- Purchasing practices
- Existing capital investments
- Difficult competitive environment
- Contractors with volatile work forces
- Builders with work forces that are difficult to manage or control

Examples of changes to the sociocultural environment that may facilitate change:

- Creation of new topic committees in professional organizations or creation of new centers or curricula in colleges and universities
- Changes to metering rules
- Changes to state and local codes
- New services for manufacturing extension services
- Simplification of product lines
- Changes in stocking practices
- Energy efficiency or clean energy as a value in corporate culture

Metrics that track changes in the sociocultural market environment may be valuable.

Project managers may want to consider how to track sociocultural factors and how such factors may inhibit or enhance program efforts. In evaluating the adoption of solar technology, one might want to examine what happens across different states where incentives and opportunities differ. For example, all of the following exist to varying degrees in some states and are lacking in others: metering rules, state laws minimizing local regulations that prevent use of solar technology on residences, promotion of solar by state or regional organizations, incentives, or an established solar industry.

In thinking about an evaluation, the program manager will want to consider the domain to be examined, the specific sociocultural underpinnings of

the domain, and whether sociocultural influence in other domains may need to be examined to understand the market effects and/or impacts.

The socio-cultural environment and the domains are intimately connected.

3.3.2 Personality or adopter types

Success, especially success in the early stages of a program, may result from the participation of innovators and early adopters. For instance, analysis currently underway for the commercial building program suggests that large food retailers may be one of the most innovative of the commercial submarkets (Reed, Johnson, *et. al.*, forthcoming 2007; Johnson, 2007). Even so, not all of the firms in this sector take the same approach to technology.

For example, some small- to mid-sized privately owned firms in the supermarket industry are particularly aggressive in pursuing innovation. Engineers at these firms are coming up with ideas for new technologies, e.g., for lighting, refrigeration, or systems integration and then implementing their ideas using in-house staff, vendors, or outside engineering firms. These firms are willing to take the time to tinker with the technology. These are also firms that take pride in being on the cutting edge, and they get strong backing from the owners. These firms are clearly innovators or early adopters. Other larger publicly owned firms are interested in new technologies as well and will adopt them after conducting a pilot project. However, they are often more interested in proven, turnkey technologies and less interested in technologies that may not be fully mature.

Some firms create a new innovation and others are looking for the already established turnkey version.

Companies differ in where they get information and what motivates them. Innovators and early adopters want to be on the cutting edge. They will look for ideas and are willing to take chances on their results. The early majority move forward when they see others using something new and want to emulate them.

From an evaluation standpoint, it is important to examine the characteristics of the firms and individuals being targeted to see if their characteristics

and the channels and messages from the program are appropriate.[3] Technology pilots are most appropriate for innovators and early adopters. Mature technologies and market pilots are appropriate for the early majority or as we sometimes call them, the "quick followers."

There are other characteristics that need to be examined as well. Experience in working with industrial firms shows that the availability of capital is an important determinant of whether a firm is interested in investing, particularly in energy efficiency.

In Rogers' theory, complexity is a key product characteristic that is closely tied to adopter characteristics. Bandura (2006) observes: "unless people believe that they have the efficacy to do what is needed to gain the benefits of a given innovation they have little incentive to adopt it or to stick with it in the face of difficulties. Personal self-efficacy is, therefore, one key factor governing adoption of innovations. Innovations that are difficult to understand and use are given less consideration than simpler ones (Rogers, 1995; Tornatzky and Klein,1982). Perceived difficulty is a relational attribute rather than solely and inherent one. Personal efficacy largely determines how complex things look."

3.3.3 Communication characteristics

Decide if you need to evaluate communication effectiveness.

Previously, we noted that there are two types of communication — broadcast and contagion, that is, one-to-many contacts and one-to-one contact respectively. Within the broadcast realm, there are flyers, bill inserts, television, radio, newspapers, radio, trade publications, etc. Contagion usually occurs within workgroups, neighborhoods, affinity groups, social gathering places, etc. The effectiveness of these channels is variable, and the impacts of a program may be greatly affected by the channels being used.

[3] We differentiate between a technology pilot and a demonstration. A technology pilot is used to validate a new technology. A demonstration is used by a firm to prove that a technology is appropriate in its specific situation.

The content of the message is important as well. Studies of demand-response programs indicate that messages about the incentive and the size of the incentive may be less effective in inducing participation than messages about participating for the good of the community (Baumgartner, 1985; Burby, 1985).

The Internet and websites might be considered a hybrid of broadcast and contagion. In some aspects, the World Wide Web is a broadcast-like medium in which users seek specific content. The message is fixed, and the user chooses whether to view and respond to the message. E-mail blasts and advertisements on websites are also broadcast methods. E-mail, list serves, and social networking programs, such as Facebook.com, probably more closely fit the contagion model. Research shows that EERE's counterparts in the utility sector are increasingly finding that their customers are getting information and signing up for programs on the World Wide Web. However, one program has found that if there is a delay between sign-up and implementation, the drop-out rate is higher for those who sign up through the Internet than those who talked to a person at a telephone center. It appears that the Internet channel is more anonymous and generates less commitment (Dohrmann, 2007).

Gauge the significance of the World Wide Web as an increasingly important way to distribute information and interacting with your customers and partners.

If you have access to a survey about how target audiences hear about your program, try dividing the responses into contagion and broadcast to assess their relative importance.

Many program managers and evaluators are surprised to find that contagion is at least as effective as, if not more effective than, broadcast in generating awareness and participation. Evaluation surveys usually include questions about how respondents heard about a program. The responses usually represent a mix of methods. If one divides the responses into methods of broadcast and contagion, contagion is frequently as important as or more important than broadcast responses.

The important point is that program managers need to be clear about which channels may have influenced the outcomes and whether an evaluation of the channels is important in assessing impacts.

3.3.4 Product characteristics

Contagion, which is so helpful in spreading the word about a technology, can have negative content just as easily as positive, and can have negative effects that, once in the culture can linger for years.

Product characteristics are one of the most overlooked issues in deployment. You can have a great program attempting to diffuse a poor product that results in widespread rejection, resistance, and difficulties even after a product is improved. The annals of energy efficiency are just full of such examples: reflective roofing got a bad start in the 1980s, heat pumps and heat pump water heaters got a bad reputation particularly in the Tennessee Valley Authority (TVA) service territory in the 1970s and 1980s, and CFLs initially had few characteristics that endeared them to customers. All of these instances are traceable to product characteristics, and all three continue to suffer from the legacy.

Look at the characteristics of the products and services produced by your program as well as the technologies and practices it promotes. This is absolutely essential.

Product characteristics are relative advantage, compatibility, complexity, observability, and trialability (see section 2.6). The important point is that managers and program evaluators should examine a technology or set of practices carefully before implementing a program or attempting an evaluation of a program promoting it. We include the program's products and services along with the technologies and practices that a program is promoting in this recommendation. Program managers and evaluators should list the five product characteristics, and analyze product and practices to see the fit of the product and the context in which it will be used. This is particularly important if it appears that a program or the technologies and practices it is promoting are having or had difficulty getting started in other settings.

Ask about behaviors in relation to products rather than attitudes toward products.

From an evaluation standpoint, measuring the effects of specific product characteristics can be tricky. If people buy something but don't particularly like the product, they may not want to admit their faulty judgment. They will often rate products more highly than they actually feel. Asking if they would buy another of the product, what their current use of the product is, or if they would recommend it are all good behavioral indicators of how a person or firm is prepared to behave in relation to a product.

Program managers are urged to examine aspects of relative advantage (especially relative advantage other than economic advantage), compatibility, and complexity. The following is an initial list of product specification questions.

Relative advantage

1. Does it cost more?
2. Does it change other important resource costs, such as time, organizational attention, maintenance costs, etc.?
3. Does it provide new, the same or more functions than existing products?
4. Is the functionality desirable from the user/consumer standpoint?
5. Does it have good aesthetic qualities or provide good aesthetics?
6. Does its use confer socially valued status?

Think about relative advantage beyond price.

Compatibility

1. Is it compatible with the system in which it is to be used?
2. Are the necessary complementary elements there to support the product?
3. Do installers have to be trained to install it?
4. Do installers need special tools that they do not now have?
5. Can it be installed or made to work under any condition or must it have special conditions? If so, do those conditions occur frequently?
6. Do existing systems have to be modified to use it?
7. If it doesn't fit in an existing ecosystem, is a supporting ecosystem easily created?
8. Are inputs required to make the technology or practice work and are they available?

Think compatibility with the:

- *Physical system*
- *Social system*
- *Market system*
- *User experience*

Build the evaluation accordingly.

Complexity

1. Is the function of the product easily understood?
2. Is it easy to install?
3. Is it easy to control?
4. Is the product intuitive and easy to learn to use?
5. Are users able to use the product with little instruction?

Most people and firms want convenience. Complexity is the foe of convenience – unless, of course, the complexity is cleverly hidden.

Trialability

1. Is it possible to borrow/buy one and try it?

2. Can the product be installed and used?
3. Is it possible to have demonstrations that illustrate the effect of the product?

Observability

1. Is it possible to see the product working?
2. Is there a way to dramatically demonstrate what the product does?
3. Is it possible to touch, feel, hear, or observe the result of using the product?
4. Can the results be measured and demonstrated?

There is no substitute for "touching" it, using it, and seeing what it does.

3.3.5 Diffusion stages

The stages of diffusion are helpful in thinking about the design of an evaluation. Early in a program, it is important to create awareness in target audiences and to begin to get audiences to consider the product. As the program matures, you want audiences to be deciding to adopt and implementing that decision. The diffusion stages generate a series of market effect impact questions.

- Is awareness in the target audience increasing? If not, why not? (See the later example for Energy Star.)
- Is the target audience searching for information, talking to colleagues, checking out information at professional meetings, etc.? Is the target customer contemplating pilots or trying a technology or practice? Is this increasing over time? If not, why not?
- Have members of the audience decided to try the technology or practice but not implemented yet?
- Have members of the target audience actually used a practice or technology? Are these numbers increasing over time?

One way to use the stages is to see how the target audience is moving through the stages (see section 2.1). A program could be having a substantial impact in terms of awareness, persuasion, and decision making, but implementation may not yet have occurred. In some instances, it may take months or years to move along the diffusion track. This suggests that a program manager and stake holders need to consider how long it will take for

a technology or practice to be adopted and to be patient. A manager in an industrial plant or chief engineer in a commercial building may decide to upgrade the efficiency of the process or building but may have to wait for a budget cycle or two for capital to be allocated.

The other side of this is that a program may get stuck at a particular point along the way. There may be a lot of interest in a technology or practice, but it does not appear to be moving forward. As an example, there is currently high awareness of refrigerator case LED lighting among supermarket energy engineers (Reed, Johnson, *et.al.*, forthcoming 2007; Johnson, 2007). Most are quite knowledgeable about it, and some have actually conducted pilot demonstrations, but many are sitting on the sidelines waiting. Many have mostly made the decision to adopt for maintenance cost reasons rather than energy cost savings but have not implemented the decision because the product cost is high and there are still issues with the technology (Taylor, 2007).

If people aren't moving through stages of adoption over time, it is likely the buzz isn't there. Check on communications and your product, especially your product.

The bottom line for program evaluators is that managers need to understand where they are in the diffusion process and what the implications are for their program's ability to move forward.

3.3.6 Replication, emulation, and sustained behaviors

Confirmation and the allied processes of emulation, spillover, and sustained behaviors should be a principal focus of evaluations. For many programs, the largest effects will not be the direct effects from installation of technology or the use of a practice, but the additive and multiplicative effects of:

Large savings or impacts for most EERE programs are likely to come from replication, emulation, spillover, and sustained behaviors. Be sure the impact evaluation focuses on these issues.

- Replication — adopters repeating the use of a technology or practice in similar situations, for example, a building contractor installing a radiant barrier in additional dwellings after first installing it.
- Emulation — non-adopters observing the use of a technology or practice and deciding to try it, for example, a competitor of the building contractor observing the first contractor installing a radiant barrier and deciding to find out about it.

- Spillover — an adopter implementing technologies or practices in addition to the ones the adopter set out to implement, for example, deciding to install an attic fan in addition to adding a radiant barrier.
- Sustained behaviors — an adopter repeating a behavior in the future, for example, a facility engineer replacing a failed CFL with a CFL rather than reverting to an incandescent lamp. Another example would be a firm creating a standard or purchasing policy specifying lighting replacements.

The key here is to make sure that everything that should be is counted and that non-program-induced impacts are not. Some of the impacts cause things to happen that wouldn't have happened otherwise, and other impacts cause things that would have happened anyway to be accelerated. Both of these effects need to be addressed.

Generally, the replication issue can be addressed straightforwardly in surveys. One can ask if the technology or practice has been duplicated elsewhere at the same site or at other sites. If replication has occurred, then additional questions may be needed to determine the extent and impact of the replication.

Likewise, follow-up surveys or follow-up inspections can address the issues of sustainability. The issues to be addressed here have to do with whether the technology or practice is still in use, whether it is being maintained, and whether the technology or practices will continue to be used.

The emulation issue is much more difficult. The question is whether other customers decide to use the technology or practice as a result of seeing a program participant using it. The original adopter may tell others about it. A contractor, distributor, or installer who worked for the initial adopter may tell someone about it. One can ask recent adopters how they heard about the technology to see if they reference earlier adopters who heard about it from a program. The data might be collected and analyzed using tools and techniques for network analysis. Samples of non-program participants may say they heard about a technology or practice by word of mouth. The question is, "was the ultimate source of the word-of-mouth information generated by the

program?" One can ask, but tracking this kind of information to its source and relating it to program participation is difficult.

Another approach is to ask program participants with whom they may have shared information and by what methods. These respondents may know if those with whom they shared information were operating in a professional capacity and how they may have used the information. One might conduct a follow-up survey with the secondary adopters.

3.4 Module 4: Identify Researchable Questions and Metrics

We will not pursue a lengthy discussion of researchable questions and metrics here. Chapter 5 provides a more in-depth presentation of these. It is important to identify questions that the evaluation will address and to identify the specific metrics. In a larger context, researchable questions and metrics should be developed with the following things in mind:

Metrics keep you on track, contribute to your reporting requirements, help define future evaluations, and help sustain programs.

- Some metrics should contribute directly or nearly directly to the organization or government agency (e.g., DOE) reporting requirements.
- Metrics should be developed with an eye to future evaluations. One of the significant problems with evaluations is that questions and metrics often shift from evaluation to evaluation. Consistency of measurement is an important aid to establishing long term trends and is, therefore, important.
- Ultimately, it will be useful to report common metrics from evaluations across all deployment programs in an organization or government agency. Program managers and evaluators may want to observe what other programs are doing, so that there is some commonality of metrics across similar program activities (such as training).

3.5 Module 5: Study Design

This is an extremely important topic for impact analysis. Without proper study design it can be very difficult to attribute effects to a program. Because of the importance of this topic, a more detailed discussion is offered in Chapter 6. This topic is also discussed in the *EERE Guide for Managing General Program Evaluation Studies*. The most important point is to design evaluation studies so that they are:

After understanding what our customers do in response to our programs, probably the next most important thing in an impact evaluation is getting the evaluation design right.

- Completed independently of the program by an objective party.
- Defensible.
- Designed to show that the effects and impacts are attributable to the program.

3.6 Module 6: Conduct the Evaluation Study

For specifics on conducting an evaluation see the EERE Guide for Managing General Program Evaluation Studies.

There are several important pieces of advice with respect to conducting the evaluation study. Some of this is discussed in more detail in the *EERE Guide for Managing General Program Evaluation Studies*.

The basic principles are as follows:

1. Assign someone to serve as project manager with responsibility for the day-to-day management of the evaluation. This person should have credibility with the program as well as the authority to manage and facilitate requests from evaluators to the program managers. A common mistake is to place a very junior staff member (rather than a very senior manager) in charge.

Use reviewers often and wisely.

2. Form a committee to review the content of the Request for Proposals (RFP), the ratings of the proposals that are received, and the work plan and work products that are developed. Other functions of the review group should be to hold periodic progress review meetings and review work products. This committee should include users of the evaluation as well as independent experts in evaluation. The size, composition, and level of activity will vary with the size and the importance of the evaluation. The committee should have direct access to senior program managers.

A qualified contractor who is independent of the program will enhance the credibility of the findings.

3. Engage a qualified and independent contractor to conduct the evaluation. The evaluation contractor (or the evaluation contractor's firm) should not be involved in program planning or program implementation except when invited to provide insight or to monitor planning or implementation activities. The contractor should keep the sponser informed when the contractor is asked to interact with program implementers to avoid the appearance of conflicts of interest.
4. Ensure that a detailed work plan is developed based on the evaluation guidelines and the work plan is subject to external review by independent experts (see review committee, principle #2).

Quality control is a key element in any evaluation.

5. Be prepared to adjust the plan as needed because of shifts in the program or because of the realities of fielding the evaluation.
6. Insist on tight quality control. Make sure that the evaluation meets the agreed-upon plan and/or be prepared to explain why the plan had to be adapted. Both the project manager and the review committee

should pay close attention to quality control issues. For example, are sample quotas being met? Are sampling procedures being carried out in a reasonable manner?

7. Ensure that the project manager monitors progress regularly through periodic telephone calls or in-person discussions.
8. Give an early heads-up about findings to the review committee and then to the program staff.
9. Use the review committee to buffer any unwarranted demands to change the findings.
10. When the evaluation is completed, make sure that there are actionable recommendations that can be pursued and that are accepted.
11. Set aside some resources for discussions between evaluators and program implementers.

Evaluation findings are more likely to be used when audiences have some inkling of what is coming and have some opportunity to have input into how they are stated. On the other hand, the evaluator has an obligation to tell it like it is, whether the results are good or bad.

3.7 Module 7: Present the Results and Store the Data

The presentation of results is critical. In Chapter 8, we provide several examples of studies and the presentation of findings from those studies. It is important to carefully archive the report and any data sets and the documentation for those data sets that are used in the analysis. Data is retrieved and reanalyzed or used for other purposes more often than one might think.

4. Using Diffusion of Innovations to Define Outcomes and Impacts

This chapter provides a tool that should make it easier for program managers and evaluators to identify outcomes. The approach combines traditional logic modeling techniques (Jordan and Reed, 1996; McLaughlin and Jordan, 1999; Jordan et. al., 2004) with Everett Rogers' diffusion of innovations theory (Rogers, 2003).

The chapter is organized as follows:

- A generic logic model that describes deployment program delivery activities. This model illustrates the logic model concept as well as providing a frame of reference within which the theory of change is implemented.
- A generic adaptation of Roger's theory suitable for incorporation into a logic model.
- The theory of change adapted to the logic model.
- Four essential variations on the theory of change, one for each of the domains.

In this document, we refer to deployment-related outputs as **products and services**, e.g., software, publications, that are produced or used by a typical deployment program to encourage the use of energy efficient or renewable energy technologies and practices. We use the term **technologies and practices** to refer to the materials, equipment, and techniques that a program's target audiences adopt in response to the use of programs' products and services.

Products and services refer to program-generated products. Technologies and practices refer to the techniques and products adopted by the program's target audiences.

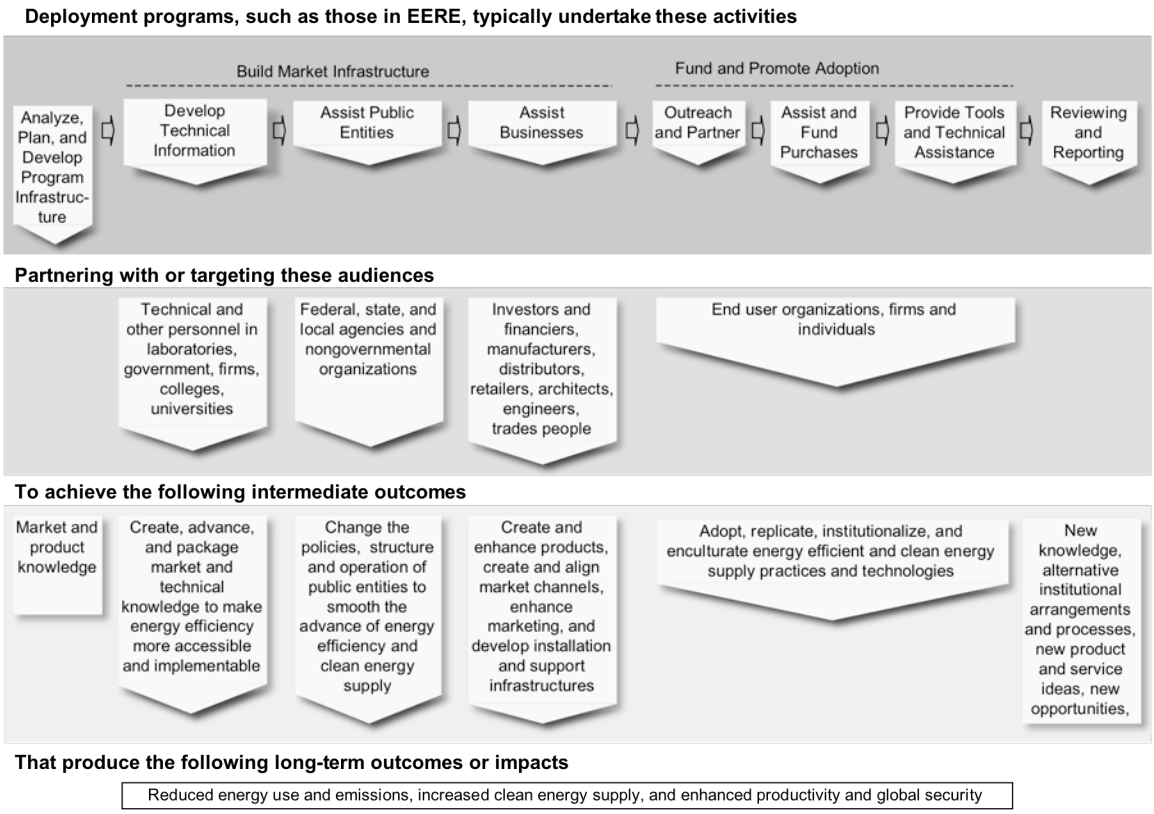
4.1 A High Level Generic Logic Model of Technology Deployment Activities

Figure 7 is a generic, high level logic model that describes how deployment programs promote the diffusion of new energy efficient technologies and practices. A separate logic model for generic research and development (R&D) activities describes the theory underlying R&D activities. That model is not presented in this document.

Use the generic model as a basis for developing a program specific model.

Multi-program organizations or agencies (such as EERE) have many different deployment programs with substantial variations in how the programs go about

doing business. At the same time, there are many commonalities across the programs. As a result, each program needs its own logic model. Individual programs can use to the generic models to facilitate the development of a specific deployment for their program (Reed, 2004b; Reed, 2005b).



Source: Reed and Jordan 2007

Figure 7. A High Level Generic Logic Model of Technology Deployment Activities[4]

The generic model presented here has four sets of high level activities: (1) analysis and planning to develop program infrastructure; (2) building market infrastructure; (3) funding and promoting adoption; and (4) reviewing, evaluating, and reporting.

All deployment programs typically have planning and analysis and infrastructure develop functions. The analysis function typically includes some market structure and market players analysis. The program infrastructure development includes developing the tools and resources that are necessary to implement the

[4] We have omitted outputs and external factors in the model to make it more parsimonious.

program. The include the development of databases, customers lists and other activities.

Building marketing infrastructure and funding and promoting adoption of energy efficient technologies and practices are at the very heart of deployment program activities. Marketing infrastructure lays the groundwork for diffusion efforts. For many Federal deployment programs, developing maket infrastructure usually entails from one to three sub-activities: (1) assembling and packaging knowledge; (2) working with regional, state, local, and public entities to establish governmental infrastructure; and (3) working with businesses to establish conditions that allow for the dissemination of technologies and practices. Thus, programs:

Developing infrastructure entails assembling and packaging knowledge, developing public infrastructure, assisting in establishing business infrastructure, and targeting end-users.

- *Identify, assemble, organize, and package technical knowledge that provides the foundation for the promotion of technologies and practices.* Most programs spend money and staff time developing outputs such as case studies, websites, manuals, software, brochures, and other materials that can be used to analyze the potential for applying technologies, explaining the value of technologies, and showing or demonstrating how technologies can be implemented.
- *Assist in developing governmental infrastructure that is essential for disseminating and promoting programs.* For example, programs may assist states and localities by providing training that helps states to establish standards or training on emerging standards such as building codes. Programs may establish public and private partnerships that can help to introduce technologies such as alternative fueled vehicles. Programs may provide information that leads to the development of utility or state standards for interconnections and net metering that can accelerate the adoption of renewable technologies such as wind and solar.
- *Assist in establishing business and infrastructure that is essential to the diffusion of technology.* This might range from setting standards for labeling (e.g., Energy Star) to training consultants and vendors in the use and application of energy-efficient and clean energy technology.

Funding and promoting adoption involves outreach and partners, assisting and funding purchases, and providing training and technical support.

Another key activity is promoting and funding the adoption of technology. As with developing infrastructure, deployment programs typically engage in three sets of sub-activities: (1) outreach and partner with organizations; (2) assist and fund purchases; and (3) provide training and technical support, as discussed below.

- *Outreach and partner with organizations.* There is a wide diversity of outreach mechanisms that include: websites where users can download fact sheets, case studies, and software tools; technical information call centers (e.g. Washington State University's Call Center for EERE that answers user questions and provides information about technologies and practices); the distribution of publications and CD-ROMS; mailings; and newsletters.

  The ITDP has an Allied Partner program that leverages the resources of utilities, manufacturing firms, and others to promote energy-efficient industrial practices through training and other means.

- *Assist and fund purchases.* DOE's Federal Energy Management Program (FEMP) has helped Federal agencies to improve the energy efficiency of Federal buildings through a performance-contracting program. FEMP also assists Federal agencies to establish purchasing practices that promote energy efficiency of products. The ITDP has used a solicitation process with manufacturers to fund demonstrations of advanced energy-efficient industrial technologies. DOE's Weatherization Assistance Program provides grants to states, which in turn contract for services to weatherize low-income homes.

- *Provide training and technical support.* Training and technical support are common activities in government deployment programs, such as those within EERE. The ITDP conducts numerous courses to train users on the software tools it has developed. The same program provides training for consultants who work in industry and recognizes those specialists on their website. Similarly, FEMP provides training for the tools it provides to its target audience. Technical support comes in many forms. The ITDP provides Plant Wide Assessments (now referred to as "Energy Saving Assessments") to aid industrial customers in improving the energy efficiency of their plants. FEMP provides technical assistance to Federal agencies to design more energy-efficient buildings.

All deployment programs have review and reporting functions for internal and external reporting requirements. As noted elsewhere, performance reporting is increasingly important and is being given additional attention. Review and reporting provide feedback to the earlier activities in terms of gaps in knowledge, the role and effectiveness of governmental and nongovernmental partners and the need for additional efforts from them, the capabilities of manufacturing and business partners and the need to augment those capabilities, and the effectiveness of program activities and the need to improve their effectiveness.

4.1.1 Target audiences and partners

In the generic logic model, deployment activities are linked to target audiences and partners. In the case of EERE, developing technical information is linked to technical and other personnel in laboratories, government, firms, colleges, and universities. For EERE, activities associated with assisting public entities target Federal, state, and local governments, regional governmental agencies and partnerships, nongovernmental organizations, and utilities. Activities associated with business infrastructure target manufacturers, distributors, retailers, consultants, and other commercial firms. The target audiences for funding and promoting adoption are firms and individuals.

4.1.2 The outcomes

Deployment activities motivate the target audiences and partners to engage in behaviors that produce results that meet a program's overall long term goal. In general, deployment activities encourage partners and target audiences to:

- Create, advance, package, and market technical knowledge to make energy efficiency and renewable energy accessible and implementable.
- Change the policies, structure, and operation of public entities to facilitate the advance of energy efficiency and renewable energy.
- Create and enhance products, create and align market channels, enhance marketing, and develop installation and support infrastructures.
- Adopt, replicate, and institutionalize energy efficient / renewable energy technologies and practices.

For a program to establish that it is causing the desired effects, it must establish the causal linkage between the outputs of its programs and the actions of its partners and target audiences. It must also demonstrate that the actions of the partners and target audiences result in outcomes that produce its long term goals.

The evaluation must demonstrate the linkages between outputs and what program partners and target audiences do (outcomes).

Deployment programs must rule out alternative explanations for target audience behaviors. To do this, the program must understand how its partners and target audiences behave and how its actions influence those behaviors. In effect, deployment programs need to describe a theory of change and then document that actual events correspond to the theory. Using the theory, program evaluations can

Understanding what partners and target audiences do can be aided by using the theory of diffusion of innovations.

identify, measure, and track outcomes, so that the linkages between outputs and impacts are apparent, and market effects and impacts can be accurately attributed to the program.

The description of activities can be facilitated by incorporating Rogers' theory of change described in Chapter 2 (i.e., diffusion of innovations) into the logic model. Rogers' theory of change provides a set of hypothesized concepts and linkages that can be used as a guide when attempting to describe how outputs generate the desired long term results.

4.2 Application of Diffusion of Innovations to Defining Market Effects and Impacts

A key argument in this document is that the effects of a program can be made more transparent and identified more systematically if one uses a theory to identify program outcomes and then designs the impact evaluation to identify and measure those outcomes. Figure 8 represents an adaptation of the diffusion of innovations model that was presented in Chapter 2 that can be used to link program outputs, program outcomes, and impacts (Reed and Jordan, 2005b; Reed, 2006b; Vine et. al., 2006).

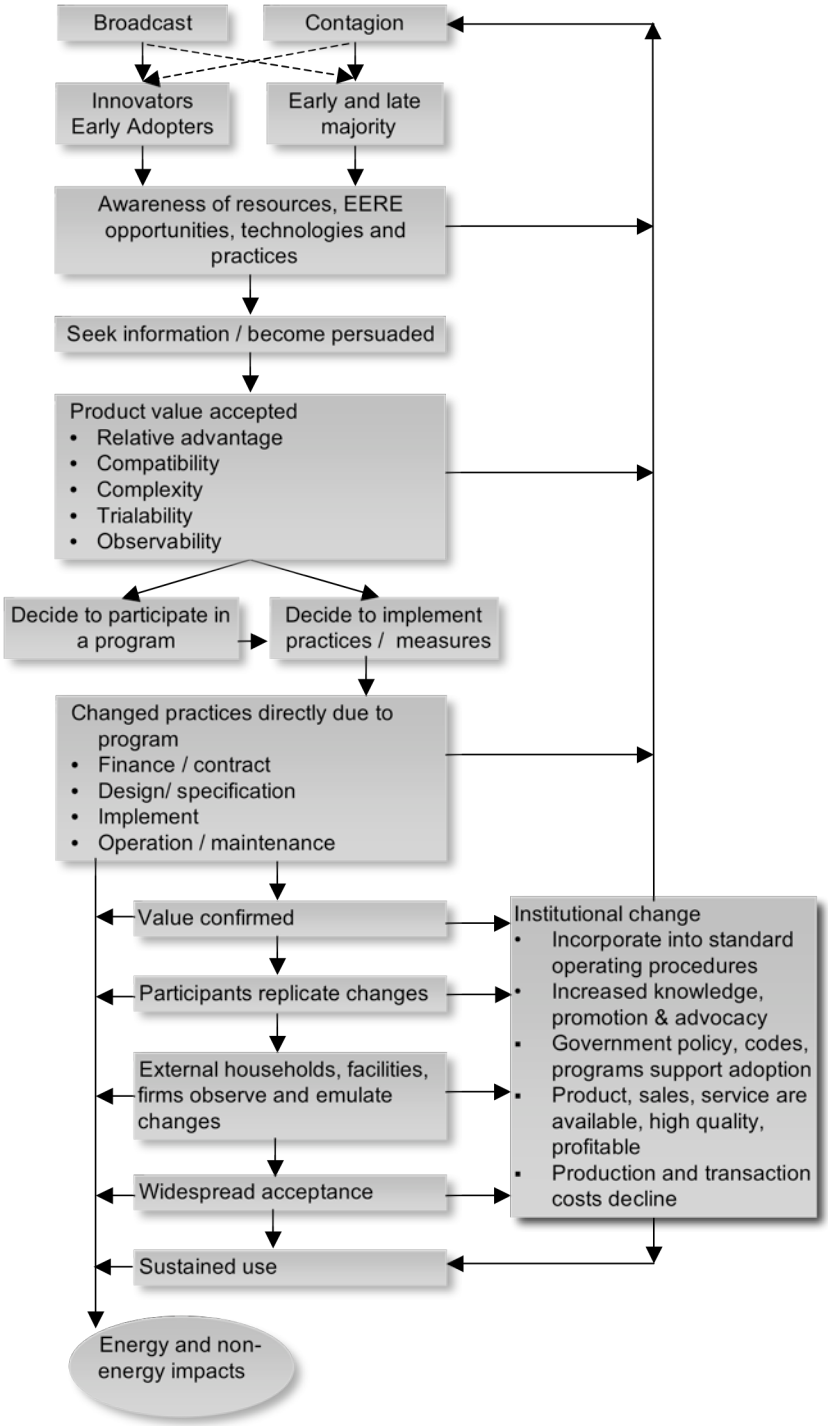
Rogers' model can be created in a form specific to a given deployment program.

Starting in the upper left portion, the target audience receives information through broadcast and contagion processes. As noted in our discussion of the theory, innovators and early adopters are more likely to get information through broadcast methods while the early and late majorities are more likely to get information through contagion processes. Dotted lines are used to remind us that the correlation between communication processes and adopter types is not perfect.

There is a dual persuasion process. A program or its partners have to sell their "wares" as well as the technologies and practices

As a result of the communication, members of the target audience become aware of the program, its products and services, and the technologies and practices the program is promoting. Some subset of the target audiences will find the information salient, and some of these may seek additional information or be sufficiently convinced that little additional information is needed in order for them to make a decision.

There is a dual information and persuasion process occurring. Members of the target audience may be seeking and evaluating information about the *products and services* provided by the program, and they may be seeking and evaluating



Source: John H. Reed and Gretchen Jordan 2005

Figure 8. An Adaptation of the Diffusion of Innovations Model

information about the *technologies and practices* that the program is promoting. Thus, a member of the target audience may find a specific technology or practice compelling but may not be interested in the program or vice-versa.

Members of the target audience evaluate both the products and services and the technologies and practices for relative advantage, compatibility, and complexity. The opportunity to try and to observe the technologies or practices will influence their decisions as well.

Members of the target audience will decide whether or not to use the program's products and services and the technology or practices the program is promoting. There are documented instances where contactors for efficiency programs have compared and updated their spreadsheets based on a tools developed by EERE, such as the Steam System tool. They have adopted the knowledge but not the tool. If the evaluator measures whether the contractor uses the tool as opposed to the knowledge, the full impacts of the tool may go unrecognized.

Actors may emulate an approach or idea but implement the idea outside of EERE's framework.

As noted earlier, implementation of the decision often lags the decision. Depending on the targeted market actor, the form of the implementation action can vary considerably. We have identified five underlying types of implementation actions:

1. Budgeting, funding, and/or contracting
2. Specifying and/or designing
3. Purchasing or acquiring
4. Changing operation or maintenance practices
5. Installing / using

These are generic types of action. For example, the targeted actor might: (1) arrange for capital to fund a project or enter into a contract; (2) establish the specification or design; (3) purchase or acquire a new technology; (4) change maintenance or operating practices; or (5) install and use a technology. Similarly, public entities might set a budget or fund an activity, establish legal requirements

or standards, establish administrative mechanisms, adjust operations, and then implement actions.

Frequently, there is a sequence of actions that flows from the decision to implement. A facility manager may recommend to a Chief Financial Officer (CFO) that a project be done, and the CFO may recommend to the Board of Directors that capital be set aside to fund the project. The Board of Directors may approve the capital request. Architects and engineers may be engaged to design the project, etc.

Implementation usually involves a sequence of actions with multiple actors over time.

On the other hand, a facility engineer may be able to change maintenance practices or operational schedules immediately. A well thought out theory can be used to identify a plausible chain(s) of events.

Confirmation occurs after a project is implemented. Like implementation, confirmation can take many forms. It may range from the impressionistic (someone casually glancing at an energy bill to confirm that the bill has gone down) to the formal (a rigorous metering and verification study with before and after metering and analysis to confirm that energy consumption has changed). A specific example of confirmation is the multifamily operator who, having installed a new type of boiler, requests a savings-analysis to determine if he should install this new type of boiler at other multifamily locations that he owns.

People may not be very sophisticated about savings and mistake externally induced effects, such as the effects of changes to energy consumption due to changes in energy cost or temperature, as being indicative of the effects of implementing an action. For example, more than a few audit customers have concluded that an energy audit either worked or didn't work in the months following the audit, because they saw their energy bills increase or decrease due to the weather or changes in the price of energy, rather than the effects of the measures installed.

People tend to see the impacts that they want or expect to see. Often the impacts are measurable but sometimes they are not.

Generally, there is a bias in the direction of perceiving positive results from having acted. For example, the project managers for a load management project declared the project a smashing success only to discover later that there was no

effect because the control equipment failed during the experiment. Mistaken perceptions can (and often do) reinforce or undermine the decision to implement a technology or practice.

As noted earlier, experiences with CFLs, cool roofs, and heat pumps led to a discontinuation or skepticism about the technologies among end-users. While some innovations have a less than illustrious history with respect to repeat purchases, other innovations find acceptance. For example, the ITDP has reported numerous instances where industrial facilities have used ITDP tools to examine a system in a facility and then repeated the use of the tool at the same or other facilities owned by the company.

It is the trying, the confirming, the replication, the emulation, and the sustaining of behaviors that are important. Unlike the innovators and early adopters, the early majority has less tolerance for "flaky" products. The early majority adopts by emulating their peers, and this is the key to widespread acceptance of a technology. An innovation accepted by the early majority will diffuse to other end-users, manufacturers, retailers, consultants, builders, and households. The early majority take their cues from each other and receive information through word-of-mouth.

For some deployment programs, the replication, emulation, and sustainability effects are likely to be many times the direct effects. That is why identifying them is so important.

The rest of the model shown in Figure 8 focuses on sustained institutional change and is an extension of Rogers' work.[5] Sustainability is not a matter of just persuading more and more decision makers to adopt a technology or practice. It involves inducing long-lasting attitudinal, structural, and cultural change. If one is targeting commercial buildings, it is not sufficient to get a firm to adopt energy-efficient lighting technology when it builds new buildings or has tenant changes. Such a policy may change when the champion of the policy leaves her

[5] The index in the Fifth Edition of Rogers' book (Rogers 2003) lists just two pages on which the word "sustainability" is used. In one case, Rogers discusses how innovations may be "re-invented" and that re-invention leads to faster adoption and a higher degree of sustainability. The idea of re-invention is that the innovation is not invariant but modified slightly in ways that might make it more acceptable. The other mention of sustainability occurs during a discussion of change agents and how change agents relate to target audiences. An innovation is more likely to be sustained if change agents are empathetic with their target audiences. Although Rogers was well aware of the importance of maintenance of the decision to adopt, he did not use the concept of sustainability.

or his position. Thus, the model for change anticipates a series of complementary changes, such as changes to standards and operating procedures, increased knowledge and advocacy, support for governmental changes to codes and standards, improvements in the quality and reliability of products, reduced manufacturing costs, and increased profits. As an example, evidence that a retail grocery firm has truly embraced the culture of energy efficiency might include hiring a full-time energy manager, establishing a corporate energy efficiency goal, tying part of the annual performance bonus of store managers and refrigerator technicians to improved energy efficiency, etc.

4.3 Integrating the Adapted Model of Change with the Logic Model

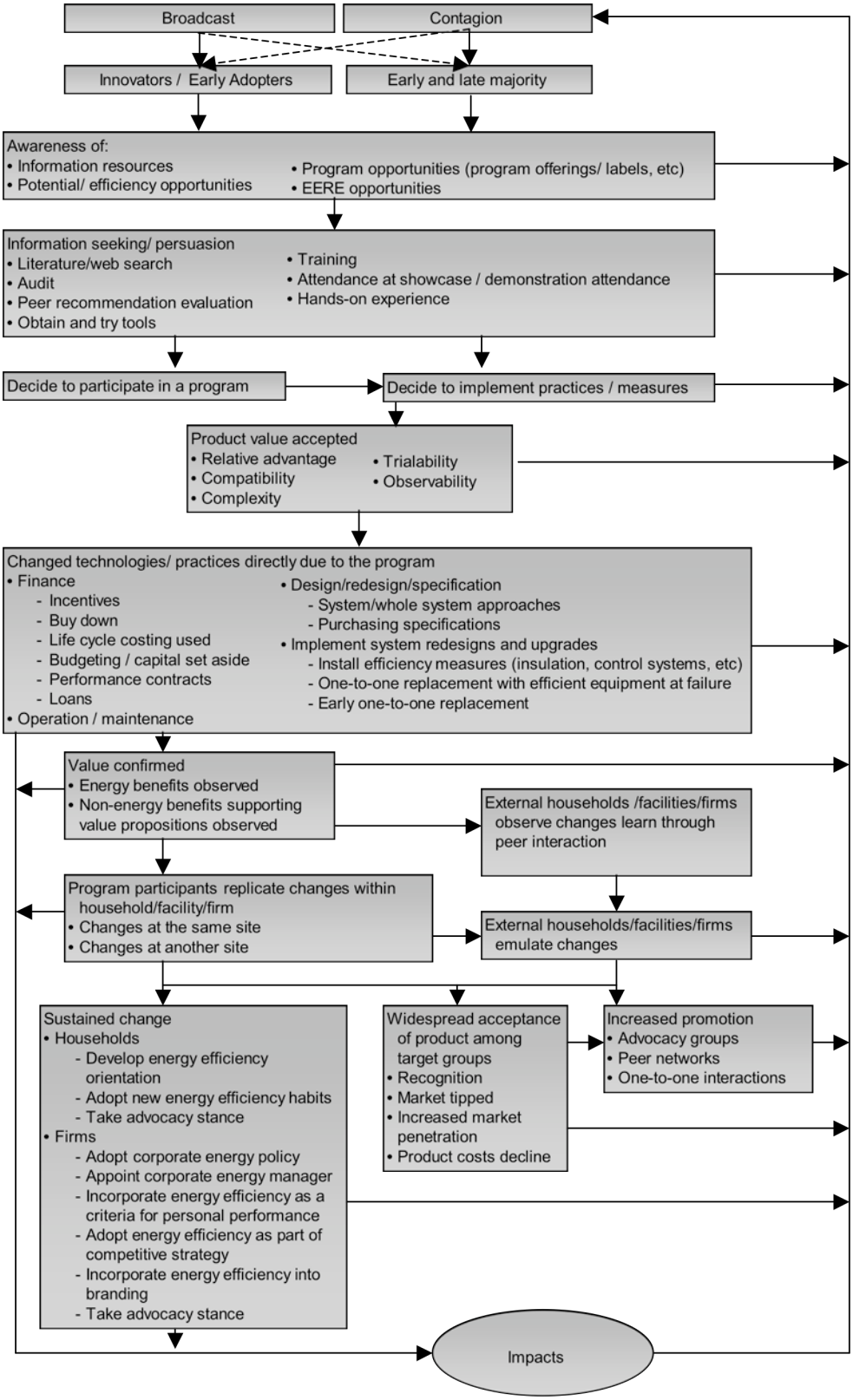
For the deployment programs generally and for programs like those in EERE, a theory is needed that is tuned to each of the major domains that the programs influence: the knowledge sector, public entities (e.g., government and nongovernmental organizations), business entities (e.g., manufacturers and other businesses), and end-users. Using the basic model of change, a version of the theory that is tuned to each domain is introduced. In the following pages, the model is applied to the four domains of interest. We start with the end-user domain because that is the domain that is most familiar to most program managers and may be the easiest to understand.

A more specific version of the Rogers' model has been developed for each of the domains.

4.4 Application to the End-user Domain

Figure 9 is a variant of the basic model describing the theory of change as it applies to end-users (Reed and Jordan, 2005b; Reed, 2006b).

A key difference between this and the generic model is that the model has been adjusted to reflect how end-users (e.g., individuals and firms) might react to a program. If the program evaluation is to focus on end-users, you should examine this model and apply it to your program to see if you have captured most of the outcomes. What follows is a brief walk through the model to see how it applies to end-users.



Source: John Reed and Gretchen Jordan, 2005

Figure 9. Diffusion of Innovations Applied to End-user Outcomes

The term end-user describes a relationship and is not a fixed attribute of a person or organization. For example, a local government may adopt codes and standards in their policy setting role and be an end-user when implementing energy efficiency programs that affect buildings, vehicular fleets, etc.

With respect to awareness, end-users may become aware of specific programs (e.g., EERE's FEMP, ITDP, and Weatherization Assistance), or a specific kind of energy efficiency opportunity (e.g., energy-efficient appliances or efficient lighting). This may lead end-users to seek information from a variety of sources including training programs, the Web, audit programs, etc. Some of these sources may be program-generated, while others may be external to the program. To identify these sources, it is necessary to examine awareness and sources of awareness among end-users.

If there are no data to support the hypothesized linkages, then it is likely that there are no effects.

This generic model does not specify a particular technology or practice. It is up to the program manager or evaluator to identify the specific technologies and practices of interest. The model shows that the end-user makes an assessment of the technology or practices by assessing one or more aspects of relative advantage, compatibility with the end-user's environment, and complexity. The end-user's decisions will also be informed by the degree to which it is possible to try the technology or practice or observe the effects. From the perspective of the impact evaluation, we want to know what criteria are being applied and with what result.

Program managers need to specify the particular technologies and practices to which the end-user model applies.

For end-users, the actual decisions made in response to information and persuasion can vary widely. An industrial end-user may decide to redesign a process or replace a motor. A residential customer might choose to install an energy efficient appliance. The decision stage is about deciding to do something, not implementing the decisions. From an impact perspective, the question is whether the decision to implement is being made, and if not, why not?

At some point after deciding, end-users targeted by the program may also implement decisions in different ways. As noted earlier, implementation may be a single step or a multi-step process that requires a series of actions ranging, for

example, from obtaining capital through actually constructing an energy-efficient building and commissioning it.

From an evaluation perspective, it is important to be able to link (or attribute) the end-user's actions to the program. If someone from an industrial firm attends a program-sponsored steam system training session, it is important to determine whether the decision to upgrade the system was made prior to the training or as a result of attending. If someone from a firm takes the training course, the firm may take action sooner than might have been the case and/or the firm may obtain greater savings than would have been the case because of the training. These are difficult questions to answer, but they are important in determining impacts. If there is no direct or indirect link to a program, then one cannot logically argue that the program influenced the choice to upgrade the system.

Once the decision is made and implemented, the actor confirms the value of the decision and the value of the implementation. For end-users, there may be both energy and non-energy benefits (e.g., convenience or comfort). An industrial firm that retrofitted it's silicon production furnaces to make them more efficient not only reduced energy use but also reduced cycle times, use of inert gas, materials residue (waste), improved the effectiveness of the product, and reduced the number of defective products (Reed, 1999a). From an end-user perspective, the non-energy benefits were substantial and valuable, perhaps more so than the energy benefits. The rest of the model suggests how one change can cause complementary changes and how the changes may become sustained over time.

For program managers who have programs that target end-users, the end-user model can be used as a template to begin to describe a more specific logic model for their end-users. A program manager can walk step by step through the model and begin to describe the actions that one might expect of end-users.

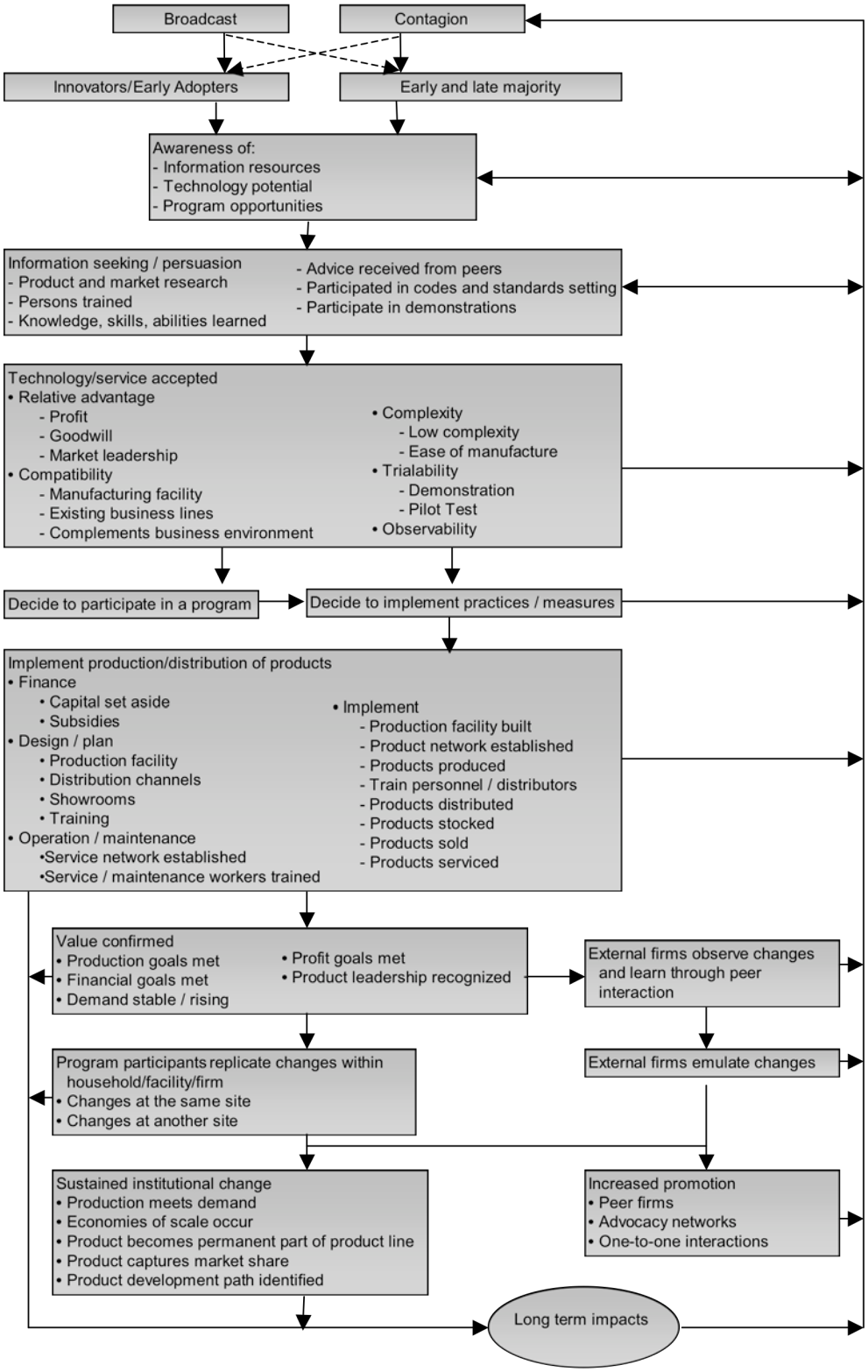
4.5 Application to the Business Entities Domain

Figure 10 shows how the theory of change model has been applied to business entities (Reed and Jordan, 2005b; Reed, 2006b; Vine et. al., 2006). When writing about the business entities domain, the discussion is about manufacturers and other businesses as suppliers and intermediaries for efficient goods and services and not as end-users. An example of the difference is a company such as Best Buy selling and promoting energy efficiency products (business entity) as opposed to increasing the energy efficiency of a Best Buy store (end-user). A manufacturer may use the ITDP program to make the plant more efficient (end-user) or manufacture efficient products (business entity).

For the business domain, the theory of change has been adapted to capture business related examples of awareness, information seeking, the application of the product model, the changes that might be implemented, and the processes of confirming and sustaining change. While the underlying theory is the same, the application environment has shifted.

Awareness of the potential in making or distributing energy-efficient technology can come from numerous places. A manufacturer or other business might become aware of an opportunity as a result of discussing or examining information about technology including its own in-house research. A business or manufacturer might be exposed to technologies from research or similar programs. Reading trade publications or attendance at trade shows might trigger awareness. Interest groups promoting efficiency equipment might contact businesses about efficient products. A trade association may contact manufacturers or retailers about proposed codes and standards efforts. From an evaluation point of view, we would want to know how widespread awareness is and how it is changing.

In response to this awareness, the manufacturer or other business may begin product and market research exploring the advantages and disadvantages of the technology or practice especially in relation to technologies and practices currently offered. Does it have relative advantage in terms of profit, good will,



Source: John Reed and Gretchen Jordan, 2005.

Figure 10. Generic Version of Diffusion Model Appropriate to Manufacturers and Other Businesses

market leadership, and other valued characteristics? Is the technology or practice compatible with manufacturing facilities, existing business lines, the availability of knowledge, and skills? Can it be handled through existing distribution systems, or will it require its own system? Would the technology add something to the business's or manufacturer's product line? Would the product compete with or undermine a current product line? The business is likely to evaluate whether the product will provide some form of competitive advantage. The cost of manufacturing or stocking the product might be explored. The firm might examine the human resource requirements imposed by the product including knowledge, skills, and abilities to make, sell, or install the product. A firm is likely to consult peers. The firm might produce a few prototypes or might conduct focus groups to examine the market. The firm might get involved in codes and standards setting as a way of obtaining information (ADM, 2006).

In the end, the firm is going to make at least some initial assessment of the potential for the technology or practice. If that assessment is positive, the firm may commit to an initial foray into the market with the product. From an impact assessment point of view, it is important to understand if these things have happened or are happening.

The firm may examine the complexity of the technology or practice from a number of perspectives. Is it difficult to build? Is it difficult for customers to use? Will customers be put off by the complexity?

As noted above, the business or firm may decide to do a demonstration or work with some selected clients to see how the technology or practice works.

Based on these and other considerations, that firm may decide to reject the technology, postpone a decision, or proceed. Previously, we discussed retail grocery firms postponing decisions about refrigerator case LED lighting. This is an example where the product characteristics are not yet right and the evaluation tells us that the product is not going to have an immediate impact. If the decision is to proceed, then firms will take steps to implement the technology or practice in the marketplace.

Implementation in the business domain involves a number of steps. The first step usually involves finance. The firm must decide how to finance the technology or pay for the initial setup of the practices. It may involve a capital set aside or budgeting money from current operations. The firm may seek subsidies or investments from outside partners.

Moving a product into the market place is a complex process.

Implementation will also require a design and / or plan. A design for production facilities may be needed. Distribution channels have to be decided upon. Showrooms may need to be designed or planned. Stock space may need to be determined. The new products may need to be added to inventory systems. Training requirements for production personnel, distributors, installers, and others may need to be planned and developed.

To implement the plans, a production facility must be built. A production line may need to be modified. Technologies must be produced. Personnel and distributors must be trained. Technologies must enter the distribution system, be stocked, and sold. Installation services may need to be provided. And once the technology is in the field, provisions for a service network and training service and maintenance workers are needed.

The important point is that moving a technology from a research laboratory to a usable product in the market place is a complex process. There are numerous places along the way where the process can become bogged down. If we are to assess the success of a program that is relying on partners to produce, promote, and sell products, we need to have a clear idea of what it takes to move a product to market. And we must be able to assess whether progress is being made in that direction. For example, many utilities want to see the movement of new efficient technologies and practices from being prototypes, to products with a share in the market, and to products that become part of codes and standards (Eilert et. al., 2002; Eilert et. al, 2004). To demonstrate the market effects of such programs, the pathways to the market must be understood, and it must be possible to trace progress along those paths.

If we are to assess the success of a program that is relying on partners to produce, promote, and sell products, we need to have a clear idea of how the product moves to the market.

It is also fairly easy to see that technologies and practices that have a high return and impose minimal disruptions to existing systems are likely to be adopted by firms sooner than technologies or practices that cause severe disruption.

In the business sector, the success of a product can be confirmed in a variety of ways. Production goals are met. Financial and profit projections are met. There is a stable and rising demand for the service or product. Peers and others recognize market leadership.

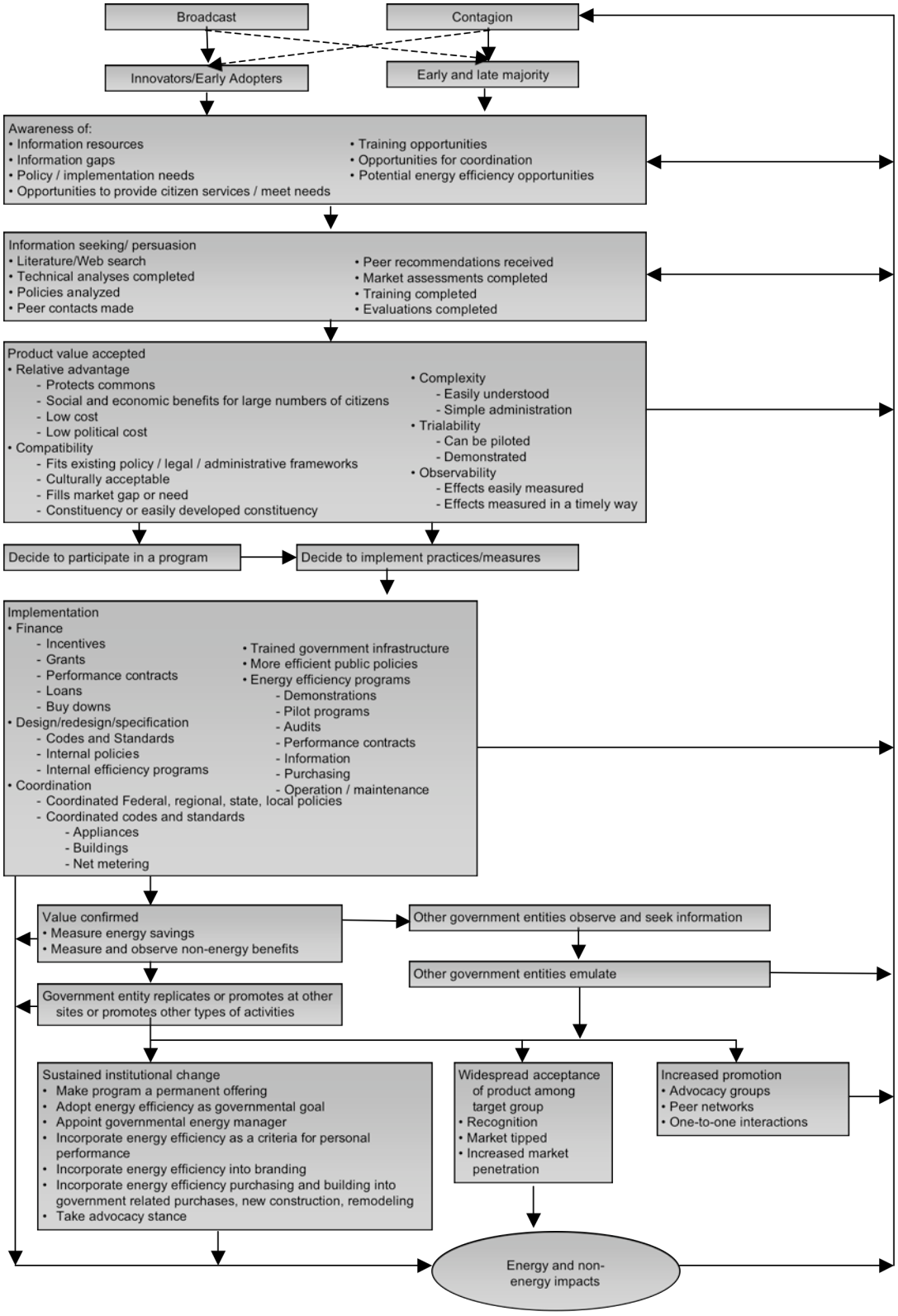
Technologies and practices can be introduced into the market, but it is also important to know if they can be sustained in the market. For a product or service to be sustainable, production needs to meet demand, economies of scale need to occur, the product needs to become a part of the product line, the product must capture market share, the product must be profitable, a product development path needs to emerge, and it may be important to create standards.

4.6 Application to the Public Entities Domain

Figure 11 is a variant of the generic model designed to show how the diffusion of innovation might apply to the public entities domain (Reed, 2005b; Reed, 2006b). As noted previously, we define the public entity domain to include governmental entities as well as a wide range of public organizations and groups that seek to promote energy efficiency or influence public policy.

Like any other entity, these governmental and public organizations have different characteristics. Some are innovators and early adopters, while others might fall within the categories of early and late majority and even the laggard. And like other organizations, different governmental and public organizations are influenced through different communications channels.

Public entities and groups have interests and awareness of situations that are different than actors in other domains. Awareness may come from the needs of citizens or it may come from programs or opportunities to meet the needs of the public that may not have been previously considered. Awareness may come from information received from other entities, e.g., effort in dealing with policy or



Source: John Reed and Gretchen Jordan, 2005.

Figure 11. Generic Version of Diffusion of Innovations Theory for Public Entities

budget needs. Awareness may also arise out of cooperation or opportunities for coordination among public organizations and entities.

Relative advantage for public entities often takes the form of protecting the commons or serving vulnerable or underserved populations.

Valued product characteristics may vary from the those valued in the business and end-user domains. For example, protecting the commons is likely to be an important value that confers relative advantage. Programs, technologies, and practices that produce social and economic benefits for larger number of citizens at low cost and at low political cost are likely to be valued. Programs and products that provide opportunities to serve vulnerable and underserved populations are likely to be important attributes.

With respect to compatibility, there is likely to be greater interest in programs and products that fit within existing policy / legal / and administrative frameworks. There is also likely to be greater attention to cultural acceptability. This may vary significantly by geography and jurisdiction. Programs, policies, and technologies that fill market gaps or needs are likely to receive greater acceptance than policies, products, or technologies that exist in (or compete with) the private sector.

Implementation mechanisms take the form of tax and other incentives, grants, and performance contracts.

Implementation is also somewhat different than in other sectors. Typical implementation mechanisms that are available to public entities are finance, design/redesign/specification, coordination, and public programs. In terms of finance, public entities can offer incentives, tax incentives, grants, performance contracts, loans, and buy downs (e.g., buying down the cost of a solar system).

With respect to design and specification, public jurisdictions set codes and standards. They can establish purchasing policies for the jurisdiction. They can establish mechanisms, such as green design programs or energy efficient housing programs in which developers may get preferential treatment for constructing green developments or energy efficient buildings.

A key role for public entities is coordination. This may include coordination between entities, consistent enforcement of building codes across local jurisdictions, or coordination of alternative vehicle fleets among different departments.

Public entities may also coordinate among groups, organizations, and interested parties within their jurisdiction (e.g., fostering support for sustainable planning or supporting a weatherization network).

Coordination is a key activity for public entities.

Public entities may also directly operate a range of programs including technology demonstrations, pilot programs, audit programs, information programs, performance contracting, and operation and maintenance programs.

Communities across the country are emulating each other's green design programs.

Communities and school districts are racing to build Leadership in Energy and Environmental Design (LEED) Certified Buildings.

Replication, emulation, spillover, and sustained change are important for this domain. Public entities may reproduce activities in other facilities that they own (e.g., auditing and retrofitting other buildings). Other public entities may observe behaviors and emulate. The benefits of audits in public buildings may lead to attempts by observes, users of the building, architects and engineers, to analyze other systems. Programs may become permanent offerings. Entities may appoint energy managers. Energy performance may be incorporated into personal performance assessments. Public entities may take an advocacy stance for energy efficiency.

Ultimately these activities lead to energy savings, reduced emissions, as well as cost savings and other effects.

4.7 Application to the Knowledge Entities Domain

Activity in the knowledge domain underlies a lot of what deployment programs like those in EERE do. Figure 13 shows how the diffusion of innovation model applies to the knowledge entities domain (Reed, 2006b). Deployment programs create, extract, organize, and produce knowledge that underpins training, software, publications, audits, energy labels, assessments, websites, and other information sharing activities.

There are numerous actors in this domain including national laboratories, universities, consultants, efficiency organizations, utilities, contractors, and public and private organization that produce research or fund research, such as the New York State Research and Development Authority (NYSERDA) and California's Public Interest Energy Research (PIER) Program. These actors own (or are aware

of) information, resources and the search skills necessary to locate and assemble information. They are also aware (or become aware) of knowledge gaps through their association with the many actors in the energy-efficiency market. Awareness may come from peers or from the serendipitous assembly of disparate facts. Program opportunities, program offerings, research announcements, and other types of needs and requests for information drive awareness of knowledge needs.

New York State's Energy Research and Development Authority and California's PIER Program are examples of entities whose mission is the creation and dissemination of knowledge.

The selection of information is subject to the same kind of persuasion / information testing process as other products. Potential users or developers of information have to become convinced about the usefulness of the information. Most of us receive and discard lots of information daily. Researchers may focus on an area where new ideas are needed, and they search for information. Information may be tested and tried.

Some utilities have research and development departments

Knowledge (like other products) has characteristics that make it more or less likely to be used. In terms of relative advantage, credibility is an important characteristic associated with knowledge. Information that comes from usually reliable sources (e.g., can be traced to known theory, or is based on data that can be replicated) is often accorded more value (has relative advantage) than information without such characteristics. The perceived economic, efficiency, production, or quality benefits associated with knowledge confer relative advantage as well. Of course, this does not mean that the information will necessarily be used to influence decisions.

Energy centers such as PG&E's Food Technology Center or the Pacific Energy Center are examples of utility organizations that are designed to create and disseminate knowledge.

One can also characterize knowledge and information in terms of compatibility. Information may stand alone. It may be consistent with existing theory or the extension of existing theory. It may fit within an existing use of knowledge (such as a piece of software) or it may require an entirely new knowledge base for its use. As an example, information may be integrated with ITDP's Steam System Tool to expand its capabilities. Or information may be created, adapted, and used to develop a tool that serves entirely different functions.

One aspect of relative advantage is the credibility of information and another is usability.

Compatibility is another key characteristic of information.

Some years ago, FEMP undertook a focus group to look at the types of tools that were needed for buildings (Reed, 2001). The consensus of the group was that

there were a large number of tools, but most were single function tools. The group determined that there was a need for a smaller set of tools whose inputs and outputs were standardized or an easy-to-use multifunctional tool.

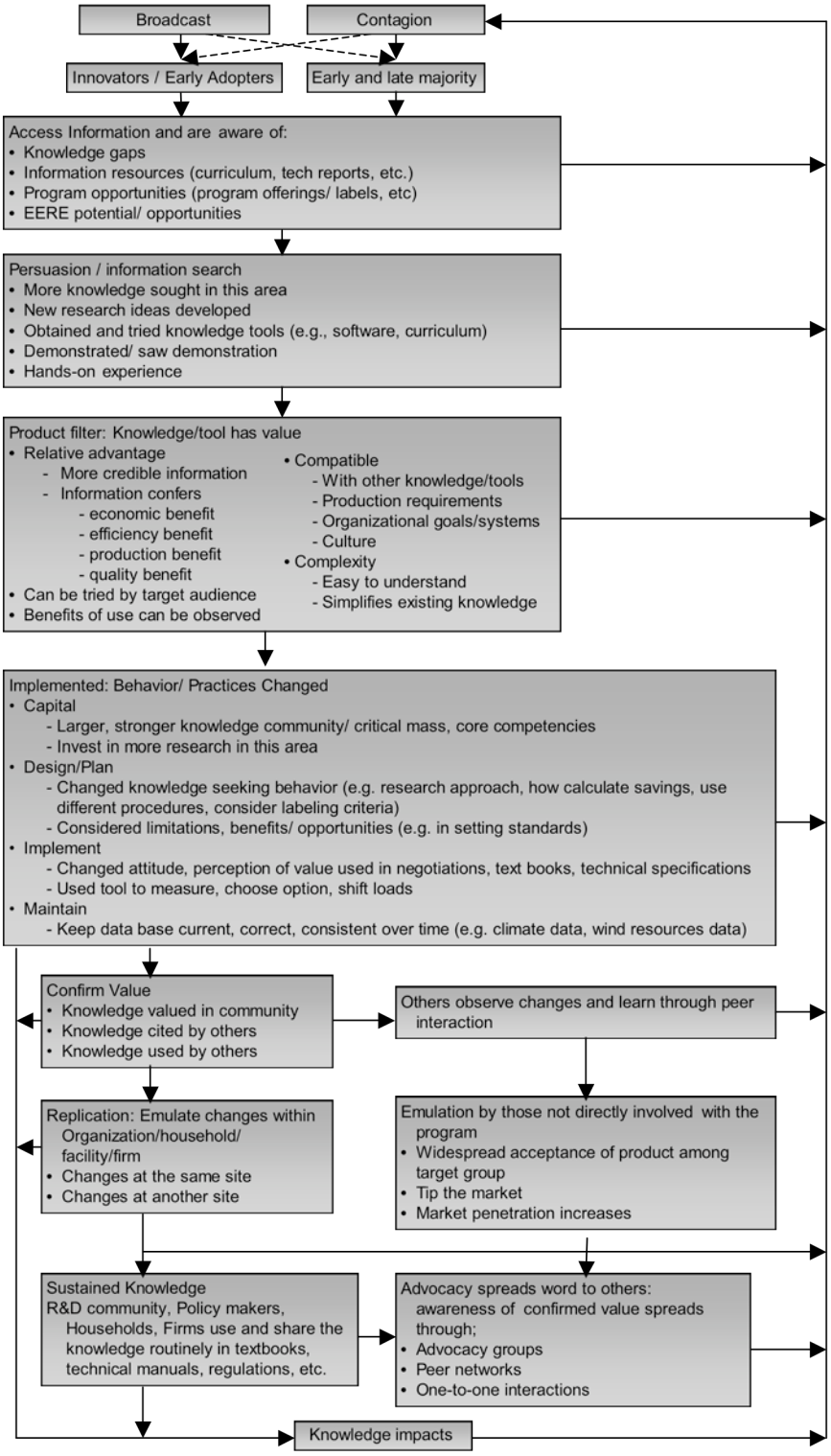
To be of value, information must be usable and people have to be able to observe that the information has value. One of the key tasks that programs perform is the reassembly of information in ways that allows people can see the value. The Energy Star labels provide information that is useable by the consumer. The observability, the low level of complexity, and the low required levels of processing are the key. That is what branding is all about, keeping it simple and keeping the overhead low.

Branding is about keeping it simple, keeping psychological overhead low, and recognition.

As in other domains, implementation involves generic activities such as investing capital, designing and planning knowledge activities, producing the knowledge, and maintaining it.

Capital investment can result in investments in new research areas or investing in existing research areas to create knowledge communities with a critical mass or a larger and stronger community with core competencies. This may be particularly important for the DOE Hydrogen Program, which is inventing new technologies that may cause significant change to existing economic and social structures (Romm, 2004). This program is attempting to increase the technical understanding of fuel cells and hydrogen.

The knowledge within much of the energy community may be more assembled than designed. For example, the knowledge within the efficiency building community has more typically grown up around specific technical issues (such as moisture control) that gained in importance and were then assembled into a program. This process is in contrast to a rationally designed plan to increase knowledge about buildings. Thus, planning and design may increase the efficiency with which knowledge is created. From a programmatic and



Source: John Reed and Gretchen Jordan, 2005.

Figure 12. Generic Version of Diffusion of Innovations Theory as Appropriate to Knowledge Entities

evaluation standpoint, it may be useful to examine the impacts of these communities and identify where gaps may exist.

Within the codes and standards processes, we see attempts to identify limitations and opportunities to formulate knowledge that can be used to set standards that raise the level of efficiency (ADM, 2006).

Knowledge is used and sustained in a variety of ways. It can be used to change perceptions and attitudes. It may be instantiated in software, publications, websites, and textbooks. It may be placed in technical specification (e.g., building guides that may be used by the architectural and engineering communities in specifying the characteristics of new buildings). The impact of these types of tools needs to be evaluated.

Finally, knowledge maintenance activities are important. It is important to keep databases updated. It is important to impose consistency in the definition of data, so that it is possible to track long term trends.

Like other domains, knowledge is subject to confirmation. Some knowledge is adopted, is highly valued, and continues to be used. Other information is found wanting and its use is discontinued. Knowledge is replicated, emulated, sustained just as with other technologies and practices.

5. Framework Questions and Metrics

Using the Rogers' model, we can generate generic researchable issues and measures that can be used by the programs to develop and check their own questions.

The generic logic model can be used to systematically identify an explicit set of impact questions and a set of metrics that flow from Rogers' theory of diffusion of innovations. Table 3 has been devised to help program managers match the diffusion of innovation concepts to researchable issues and measures.

The table is laid out in four columns. Column one presents the concept – e.g., ultimate impact or adopter characteristics. Column two identifies researchable questions that are illustrative of the impacts associated with each concept. Column three specifies illustrative impact measures for these questions. Column four identifies related process questions.

The questions are intended to make it easier for program implementers and evaluators to identify metrics for routine data collection or for periodic surveys of programs. For example, a program may want to measure the impact of a program on awareness. Going to the section on awareness, one can find examples of researchable questions and measures appropriate for market effects and impacts. With some modification, the questions and metrics can be applied across the four domains that have been identified in the previous chapter – end-users, manufacturers and other businesses, public entities, and knowledge – to generate program and domain-appropriate researchable issues and measures.

The purpose of the list is to provide a guide or reference for specific types of questions. The intent of the list is not to provide a long list of questions that the program manager or evaluator must examine. The use of the list will be bounded by the limits of the evaluation being conducted. For example, an evaluator or program manager evaluating an emerging technology may want to address product characteristics. They can go to the section on product characteristics and see some examples of such questions. Given knowledge of the specific technology, they may then generate a greater number of more specific questions.

Table 3. Framework Questions and Measures[6]

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
1. Ultimate Outcomes or Impacts	1. What were the energy and cost savings impacts? 2. What is the difference in energy used by participants before and after participating in the program (i.e., how much energy was saved)? (Gross savings) 3. What was the difference in energy use by non-participants before and after the program (i.e., how much energy did non-participants save without the program)? 4. What was the difference in the change in energy use from before to after the program for participants and non-participants? (Net savings) 5. What were the non-energy impacts (benefits / negative benefits) before and after the program?	• mmBtus • Dollars • kWh • kW • Emissions produced - tons of carbon - tons of SO_x/NO_x • Comfort experienced • Productivity - product per unit of energy	1. Compared to the program theory, are there any unanticipated impacts? If so, what are they and how large are they? 2. Does the program theory need to be modified to account for these impacts? 3. How can the program better reach and serve non-participants? Hard-to-reach populations? 4. What are participant and non-participant recommendations for enhancements to program process and content? 5. Are there areas for improvement in the program's administrative functions (e.g., marketing, recruitment, record keeping)?
2. Adopter Characteristics[7]	1. What was the distribution of the program's targeted technologies and practices[1] in the market by adopter type before and after the	• % innovators • % early adopters	1. Who is currently using the program's targeted technologies and practices? a. Innovators

[6] Throughout these questions, we differentiate between products and services on the one hand and technologies and practices on the other. Products and services refer to the materials, information media, and activities that the program produces or uses to encourage target audiences to adopt technologies and ideas. Technologies and practices refer to the innovations that the program is attempting to get target audiences to adopt.

[7] Adopter process questions 12, 13, and 14 could apply to other areas in this table as well.

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	program (or between two points in time during the program)?	• % early majority • % late majority • % laggards	b. Early adopters c. Early majority d. Late majority e. Laggards
	2. How many of each type of adopter have been reached by the program?	• # of innovators • # of early adopters • # of early majority • # of late majority • # laggards	2. What role do these participants play in the market? a. R&D community/product expert b. Policy maker c. Supply chain d. Partner/ally e. Consumer
	3. How did the number of adopter types change before and after the intervention?	• # of innovators • # of early adopters • # of early majority • # of late majority • # laggards	3. Is it easy for the members of different adopter groups to join or participate in the program? 4. What motivates members of adopter groups to participate? 5. Are program products, services, and delivery strategies consistent with adopter group motivations? 6. Do marketing materials emphasize valued benefits for members of adopter groups? 7. Which program products and services are these adopter groups using? 8. Are some program products and services under-utilized by certain adopter groups? Heavily utilized? 9. What could be done to increase the use of the program's products and services within targeted adopter groups? 10. Are some of the technologies and practices promoted by the program more heavily utilized by adopter groups? Why? (Refer to product characteristics sec-

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
			tion.) 11. What are the nature and magnitude of current barriers to the use of program-promoted technologies and practices by targeted adopter groups? 12. How can the program better reach and serve non-participants? Hard-to-reach populations? 13. What are participant and non-participant recommendations for enhancements to program process and content?* 14. Are there areas for improvement in the program's administrative functions (e.g., marketing, recruitment, record keeping)?* 15. Are the various adopter types satisfied with the program's products and services and the technologies and practices being promoted by the program?
3. **Progress of Technology through the Stages of Diffusion**	**Awareness** 1. Has the percentage of the people/firms that are aware of the program's brand increased, decreased, or remained about the same? 2. Has the percentage of people/firms aware of the program's targeted technologies or practices increased, decreased or stayed the same in response to the program? 3. How does this compare to non-participants?	• Percent of people or firms that are brand aware • Percent of people aware of the technology or practice	1. Are there aspects of the program's products and services that have been effective in increasing the awareness of the targeted technologies or practices? Are there areas for improvement? 2. Are there barriers (e.g., to awareness, information/persuasion, decision, implementation, confirmation, sustainability) that reduce awareness or participation in the program? 3. Can existing barriers be reduced or eliminated to increase awareness, information/persuasion, decision, implementation, confirmation, and sustainability?

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	Information/Persuasion 1. In response to the program, are customers and suppliers seeking more information about the targeted technologies or practices? 2. Has participant knowledge about the targeted technologies or practices increased from the period before the firm or people encountered the program or from an earlier to later time during the program? 3. How does this change in knowledge compare to that of non-participants? 4. Can the changes in knowledge about technologies and practices be attributed to the program's product and services?	• # or percent of customers seeking information about a technology or practice - on the Internet - from dealers - from friends • # or percent of customers having knowledge of product attributes – especially energy efficiency attributes	1. Are there aspects of the program that have been effective (or ineffective) in stimulating requests for information? 2. Are there aspects of the program that have been particularly effective in educating participants about the benefits of the technologies or practices?
	Decision making 1. Has the number or percent of firms, organizations, and people deciding to sell or use the targeted technologies or practices increased, decreased, or remained the same? 2. How do the numbers or percentages of firms, organizations, or people participating in the program and deciding to use technologies or practices compare to non-participants? 3. Have changes in knowledge that are attributable to the program (see information/persuasion) resulted in people or firms deciding to adopt the targeted technologies or practices?	• # or percent of persons or firms deciding to try a technology or practice • # or percent of decisions to adopt a technology or practice influenced in part or whole by program supplied information	1. Has the program helped to influence customer and supplier decisions to adopt a program's targeted technologies and practices? Are there areas for improvement? 2. Do the program's delivery strategies address customer and supplier motivation? 3. Are customers and suppliers seeking/using information about the program and about the program's targeted technologies and practices? If not, why not? 4. Does the program have elements to reduce the amount of time between decision making and imple-

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	4. Can the difference in numbers or percentages for participants and non-participants deciding to use the targeted technologies or practices be traced to other program activities?		mentation? Are there areas for improvement? 5. How long does it take for individuals and firms to make a decision?
	Implementation 1. Has the number or percent of customers and suppliers who have taken action increased, decreased, or stayed the same, in response to the program? Actions could include the following: a. Changing operations or maintenance b. Setting aside capital c. Designing something d. Purchasing something e. Installing something 2. How does that compare to non-participants? 3. Are the changes in the number of participants taking action attributable to the program? 4. For participants, has the amount of time between decision making and implementation increased, decreased, or remained the same? 5. How does that compare to non-participants?	• # or percent of participating persons or firms taking a specific action • # or percent of non-participating persons or firms taking a specific action • Energy savings per action • Actions partially or wholly attributable to the program	1. Is it easy for customers and suppliers to join or participate in the program? If not, why not? 2. What motivates customers and suppliers to: a. Participate or not participate in the program? b. Implement the program's targeted technology or practice (in addition to the product value)? 3. What accounts for different levels of customer and supplier participation in the program? 4. Are customers and suppliers satisfied with the program?
	Confirmation 1. Are participating customers and suppliers continuing to use the technologies and prac-	• # or percent of persons continuing	1. Are customers and suppliers likely to have another opportunity to implement?

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	tices after they implement the first time or do they discontinue use? 2. Are customers and suppliers replicating program induced behaviors in existing or other settings after they implement the first time? 3. How do these actions compare to non-participants? 4. Is there any delayed implementation of technologies and practices?	to use the product or service • # or percent of individuals or firms replicating behaviors • Unit savings for replicated behaviors • # or percent of adopters delaying implementation	a. If yes, are customers and suppliers using more of the program's products or services or adopting more of the targeted technologies and practices? b. If no, why not? 2. Have customers and suppliers reverted to old technologies and practices? If yes, why? 3. Are customers and suppliers satisfied with the program? 4. Are there aspects of the program that have been effective in increasing the confirmation process? Are there areas for improvement?
	Sustainability 1. After adopting, how many or what percent of individuals or firms are taking additional actions that reflect the internalization or institutionalization of the need for energy efficiency? Questions addressing internalization: a. Have efficient technologies or practices been adopted as standard by firms or individuals? b. Have there been changes in management practices as a result of the program? c. Did firms appoint an energy manager as a result of the program? d. Did firms establish energy policies as a result of the program? e. Do staff compensation packages include	• # or percent of firms internalizing or institutionalizing actions • # or percent adopting standards technologies and practices • # or percent changing operational or management practices • # or percent appointing an energy manager	1. Has the behavior (e.g., purchase and management decision making and practices) of customers and suppliers changed over the life of the program? 2. Are customers and suppliers satisfied with the program?[8] 3. Are there aspects of the program that have been effective in increasing the degree to which behaviors are sustained? Are there areas for improvement? 4. Have regulations, pricing, service, availability of financing, etc. changed due to the program?

[8] This question could be asked at all stages of diffusion.

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	components based on reductions of energy use? f. Is there other evidence of organizational culture changes? 2. Have participating customers and suppliers become advocates for energy efficiency, the program, or technologies and practices? 3. Have there been any emulation of adopted practices and technologies by other firms or organizations as a result of the program? 4. Have there been any spin-offs (new practices, technologies or additional applications) as a result of the program?	• # or percent instituting an energy policy • # or percent integrating efficiency into performance standards and measures • # or percent of firms advocating for the program, technologies, or practices • # or percent of firms emulating the program, technologies, or practices • # of spin-offs in firms implementing the program, technologies, or practices	
4. Sociocultural (Market) Environment	1. Has the number or percentage of product availability increased, decreased, or stayed the same since the program was instituted? 2. What was the saturation of the targeted technologies and practices promoted by the program in the market at the start of the program (i.e., the market baseline)? 3. What was the penetration of the technologies	• # or percent of firms or households having the technology or using the practice • # or percent of new firms or households purchasing the technology or adopting	1. Who are the key market actors and decision makers by segment and how do they interact? 2. What motivates decision makers? 3. What role do these participants play in the market? a. R&D community/production expert b. Policy maker c. Supply chain

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	and practices promoted by the program in the market before and as a result of the program (product sales)? 4. How many decision makers has the program reached? 5. Has the market or governmental infrastructure changed since the program was implemented? Have there been increases, decreases, or similar levels in: a. The knowledge base about the technology or practice, the market, or the sociocultural environment? b. Manufacturers, distributors, and retail outlets for the technologies or practices? c. Support infrastructure to support the installation and maintenance of technologies and practices? d. Policies, regulations, codes, and standards that support the use of the technologies or practices? e. The number of stakeholders or government officials – for example, code officials – who understand and can implement the code? f. The number of governmental programs that supply information or provide access to networks? g. The number of developers, manufacturers, retailers, installers, or financing specialists who know about the technologies and practices and recommend or promote them?	the practice • # or percent of manufacturers making the technology • # of efficient models available and sold • # or percent of distributors having technology in stock • # or percent of trade allies recommending or installing the technology or using the practice • # or percent of firms than can obtain installation and maintenance for technologies or practices • # or percent of state or local entities with appropriate codes and standards • # or percent of stakeholders who have been trained or are capable of implementing codes and standards	d. Partner/ally e. Consumers 4. How large is the potential market for the program's targeted technologies and practices? 5. Does the use of the program's targeted technologies and practices differ by market segment? 6. Has the program shown how the program's targeted technologies and practices are compatible with existing physical and/or sociocultural systems? If not, why not? 7. What prevents the use of the program's targeted technologies and practices?

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
		• # or percent of entities that provide information or access to information networks • # or percent of developers, manufacturers, retailers, installers, or financing specialists who know about the technologies or practices and recommend or promote them	
5. Communications	**Communication by Broadcast** 1. How many or what percent of the target audiences hear about the program's products and services from the program's information sources (e.g., media, personal, broadcast)? 2. How many or what percent of the target audiences hear about the technologies and practices promoted by the program from the program's information channels (e.g., media, personal, broadcast)? **Communication by Contagion**	• # or percent remembering or using program mailing • # or percent remembering or using program brochure • # or percent using program website	1. What are the primary methods the program uses for getting and sharing information? To what extent does it use broadcast channels (one-to-many channels) and contagion (e.g., word of mouth)? 2. Which delivery channels are working well (or not working) to achieve program's broadcast and/or contagion objectives at minimal cost? 3. How do these delivery channels operate? 4. Can the effectiveness of the message delivery channels be increased? 5. Are those who are interested in the technologies or practices interacting one-on-one with each other, or

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	1. How many or what percent of users heard about the program's product or services or the technology or practices by word of mouth (from other users)? 2. How many or what percent of the information about the program's products or services or the targeted technologies or practices is being passed on? 3. How many or what percent of users are adopting the technology or practices on the basis of other users recommendations? 4. How many or what percent of adopters are promoting the program's targeted technologies or practices in response to the program? 5. How many news articles on the program's targeted products and services have appeared?	• # or percent of adopters who tell others about the program or the technologies or practices • # or percent of adopters users who say program participants influenced their use of the technologies or practices	getting information from published sources? 6. Is there anything that prevents the communication of information about the program's products and services or the technologies and practices? 7. Have news articles been published about the program or the program's targeted products and services?
6. Product Characteristics	**General Product Characteristics**		1. Is the product more suitable for innovators and early adopters? Or is it more suitable for early majority and late adopters? 2. Are customers satisfied with the product?
	Relative Advantage 1. Are the attributes of the program's targeted technology or practice (e.g., relative price, payback, functionality, ease of use, quality,	• Price • Perceived ease of	1. Do program products or services have relative advantage? 2. Are program products or services perceived to

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	performance status) less advantageous, about the same, or more advantageous than a competing product?	use • Immediacy of payback • Functionality • Quality • Status	provide value? 3. Do program products or services provide a value that is not available elsewhere? 4. How does the program's marketing materials emphasize benefits that have high value for customers? Are there areas for improvement? 5. What does the customer perceive as the relative advantages of the program's targeted technologies and practices – including price, aesthetics, ease of use, and performance – as a result of the program? 6. How effective have the program's product and services been in providing information about the relative advantages of the program's targeted technologies and practices – including price, aesthetics, ease of use, and performance to various audiences – within the R&D community, among manufacturers, etc.? 7. Is the program examining and emphasizing product characteristics that best match consumer preferences?
	Compatibility 1. Compared to existing products, does the program's targeted product have greater compatibility, less compatibility, or about the same level of compatibility: a. With existing technology or practice delivery systems? b. With the way in which existing technologies and practices are installed or used?	• Amount of change required to delivery channels and of personnel in channels measured in dollars, knowledge, or skill • Amount of change required to existing	1. How well has the program's products and services shown how the program's targeted technologies and practices are compatible with existing physical and/or sociocultural systems? 2. If not, why not?

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
	(i.e., would the technology be installed or used in the same way or in the same system?)	systems (percent of system affected)	
	Complexity 1. Are the program's targeted technologies or practices: a. Easily installed? b. Easily used? c. Easily mastered – i.e., usable without extensive additional knowledge or learning? 2. How much has the program reduced the perceived complexity of the program's targeted technologies or practices?	• # or percent of installers or users who have the understanding required to install or use the targeted technologies or practices • Amount of user training required to install or use the targeted technologies or practices (dollars or person days) • Level of effort to use the targeted technologies or practices (dollars or person days)	1. Do the program's products and services help the adopter learn to manage the complexity of the technologies and practices? 2. Are there areas for improvement?
	Trialability 1. Can a program's targeted technology or practice be tried by a potential user with minimal hassle or risk, or is the technology or practice difficult to try?	• Degree to which user can touch and feel the product • Degree to which	1. Do products and services help the adopter try the program's targeted technologies or practices? Are there areas for improvement? 2. How much have the program's products and services

Evaluation Focus Area	Outcome or Impact Questions	Measures for Impact Questions (including units)	Process Questions
		user can control the product • Degree to which user can experiment with the product	helped to reduce the amount of learning needed to use the program's targeted technologies or practices?
	Observability 1. Can the results (e.g., operation, performance, benefits) of a program's targeted technology or practice be easily observed or is it difficult to observe?	• Degree to which the product is observ-able • Degree to which the operation or per-formance of the of the product can be sensed or observed	1. Do aspects of the program help make its targeted technologies or practices easily observed? 2. Are there areas for improvement?

6. Developing an Evaluation Design

The term "evaluation design" can carry different meanings depending on the context in which it is used. It is frequently refers to a description of the way in which methods and analysis procedures are organized to complete a study. These might more accurately be called "methods." When used in relation to an impact study, the term "evaluation design" describes the logic and procedures used to establish causality. This is the primary focus of this chapter.

For those interested in general evaluation design, the *EERE Guide for Managing General Program Evaluation Studies* contains a very useful summary of the key steps in designing an evaluation study:

Consult the EERE Guide for Managing General Program Evaluation Studies for information on populations, sampling accuracy, and analysis methods.

- Identify the questions and indicators for which data will be collected.
- Inventory the existing data and identify the data gaps.
- Identify the method and timing by which the data will be collected.
- Identify the populations from which the data will be collected.
- Make choices about research accuracy, sampling precision, and confidence level, and identify the degree of defensibility for the results.
- Identify the method of analysis used to produce the evaluation results.
- Identify a method of reasoning from results to answer questions.

Readers should consult the *Guide* for help on these issues. Two other resources are the government accountability office (GAO) report *Designing Evaluations* (GAO, 1991), especially chapters 2 and 3, and the *California Evaluation Framework* (Tecmarket Works, 2004).

The approaches to evaluation design presented in this section are focused on establishing causality or attributing effects to a program. A research design for any study based on this framework has to be customized to the particular needs of the programs, its target market, evaluation scope, and specific questions and measures.

6.1 Evaluation Design Options

First, an impact evaluation needs to demonstrate that an outcome has occurred. This requires (1) measurements of baseline or pre-program behaviors or energy consumption; (2) knowledge of program activities or interventions that would affect behavior or energy consumption; and (3) measurement of post-intervention behaviors and energy consumption within the target audience. The measurements, interventions, and predicted effects should be identifiable from program theory.

An impact evaluation must demonstrate that the impacts are due (attributable) to the program.

Second, the evaluation design must make it possible to determine that the observed changes or outcomes are due to the program and not other influences or mere coincidence. Demonstrating this attribution is usually the most difficult part of the evaluation.

Evaluators use the words "gross" and "net" impacts to distinguish non-program-induced changes in behavior and consumption from program-induced changes in consumption.

We are most interested in net rather than gross impacts.

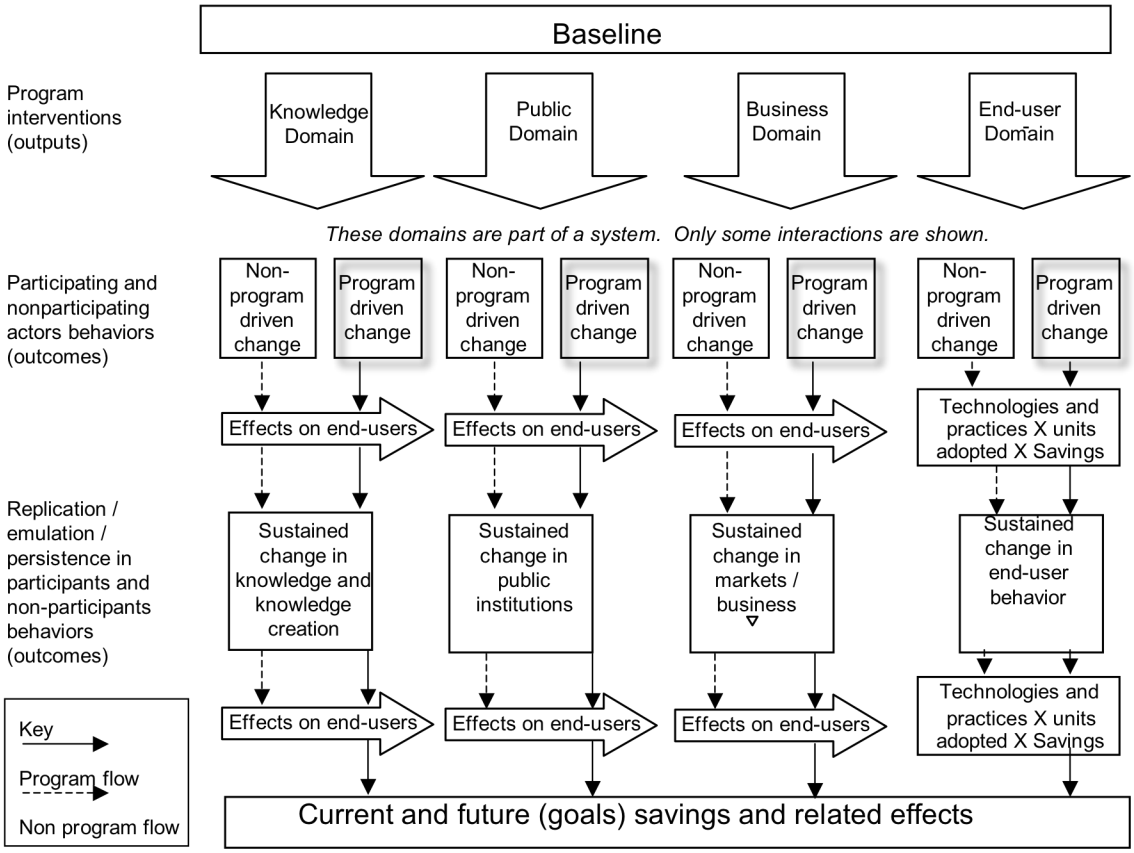
- Gross Impacts: Changes in behavior or energy use resulting from the actions of participants (whether program induced or not)
- Net Program Impacts: Changes attributable to the program (i.e., net impacts are gross impacts minus the impacts that would have occurred even if the program had not been offered)

The impact evaluation framework makes formulating evaluation questions easier because logic model templates allow organization of possible changes/effects into four domains: three related to infrastructure (knowledge, public/policy, and business infrastructure), and one related to the end-users of the technology or practice. For each of these domains that apply to a program, the evaluation should describe and assess non-program-induced and program-induced effects.

6.2 Direct and Indirect Impacts of Deployment Programs

Figure 13 shows the scope of the framework and the places where the assessment of program and non-program-induced changes need to be considered. As

noted above, program strategies for generating ultimate outcomes will typically fall into one or more of four domains. A program might be focusing on creating and packaging knowledge with special reference to the business domain. A recent example of this is a building energy program that partners with builders to produce more efficient homes. The focus is on builders who are business intermediaries in the market place.



Source: John Reed and Gretchen Jordan, 2006.

Figure 13. Program and Non-Program Effects on Actors in the Market

If we follow the business path and work down, we encounter participating and non-participating actors (in the case of the example, participating and non-participating builders). The expected program outcome is that participating builders will take action, but it is also possible that non-participating builders could take similar actions as a result of external factors.

A good example of this is two heating, ventilation, and air-conditioning (HVAC) quality installation programs. One is a statewide, residential utility-

sponsored quality HVAC training program. The second is a parallel but independent program (non-program driven change) by a leading HVAC manufacturer that operates in the same state and encourages HVAC vendors to train their installers in quality installations in return for more heavily discounted HVAC equipment. There is nothing to prohibit vendors of the HVAC manufacturer program from also participating in the utility program. Both the utility-based program and the vendor-driven program can cause quality installations. If an evaluation does not address the effects of the vendor driven program, the outcome of the utility HVAC program may be substantially overstated because installations driven by the HVAC manufacturer's program would be included in those with the utility program. Further, the manufacturers' vendors would benefit from the incentives from both programs.

It is important to figure out what factors or programs could produce the same results as the program being evaluated.

In another example of non-program driven change, interviews with some major national retailers have produced evidence that when they install equipment they check with utilities and public goods charge programs to see if their installations qualify for rebates. Indeed, there is at least one national firm with a business line that locates and notifies national retail firms about incentives and helps to collect incentives for clients that have constructed or extensively remodeled retail or office spaces. Under these circumstances, the results of the utility commercial incentive programs may be overstated because retailers and office owners might install the measures in any event. This latter phenomenon is called "free ridership."

If the incentives are included in the cost-benefit calculation and are a basis for making an energy efficient equipment choice, then the impact can be considered a program effect. If the incentives don't influence the choice or no longer influence the choice (the company has adopted a policy to install energy efficient equipment), then the incentives are simply a bonus for doing what the retailers would do anyway. There may be large commercial firms that predicate their decision to install on whether or not incentives are available. Such installations would be a result of the program but would have no replication effects.

So far, the discussion has focused on "direct" program effects – i.e., the program induced a firm or individual to do something (or to make a decision not to do something) at a given moment or in a particular situation.

In the long run, indirect effects may produce more savings than direct effects.

There are indirect or secondary effects as well. As we observed in previous chapters, the secondary effects can be greater than the primary effects. There are simply not enough resources to directly induce all the energy efficiency effects that might be desirable. Ultimately, the goal of most programs is to create enough value so that the effects are replicated, emulated, and sustained. The indirect or secondary results are from participants taking additional action beyond that stimulated directly by the program or from non-participants taking actions after observing participants.

For both theoretical and measurement reasons, we call these "replication," "emulation," and "spin-off/spillover" effects. Replication effects result from those actions firms repeat within the same facility or within similar facilities. Using the ITDP's steam tool to upgrade one steam system at a facility and then using it again on another steam system at the same plant outside the program is replication. In this example, replication occurs if the same firm decides to use the tool at another of its plants.

Are participants replicating their behaviors at other plants or other stores?

Emulation occurs when someone observes the action of others and then decides to act. An example of emulation is a competitor firm or another facility manager observing the use of the steam tool and deciding to use it as well. Spin-off/spillover refers to actions that are inspired by the program but are not necessarily the actions that a program is attempting to influence. For example, after a company makes a steam system more energy efficient, it decides to look at other systems (such as compressed air or process heat). Or the company may examine the steam system with the steam tool but, in addition, decides to improve the overall efficiency of the piping system that was not necessarily recommended by the steam tool.

Did other end-users, businesses, and public entities copy the behavior of a program participant?

This paradigm is also useful to differentiate between the concepts of persistence and sustained effects. In the energy field, persistence usually refers to

Persistence is the length of time a technology or practice continues to produce savings.

technological persistence, that is, the degree to which an energy efficiency measure continues to produce savings. Persistence is the inverse of degradation. Thus, the level of insulation (R-value) in a wall may decline over time (Figure 14), for example, from R17 to R13 over a 15-year period, and therefore, the persistence of the savings at the end of 15 years is 76 percent of the original amount.

We usually think of sustained effects as continued or repeated behaviors. Some people may refer to this as behavioral persistence. A resource acquisition program may "buy" energy efficiency. This may mean replacing a less efficient piece of equipment with a more energy efficient piece of equipment, but when the lifetime of the piece of efficient equipment expires, the actor may or may not replace (sustain) it with an efficient piece of equipment. A contractor who continues to install insulation in new homes is sustaining a behavior.

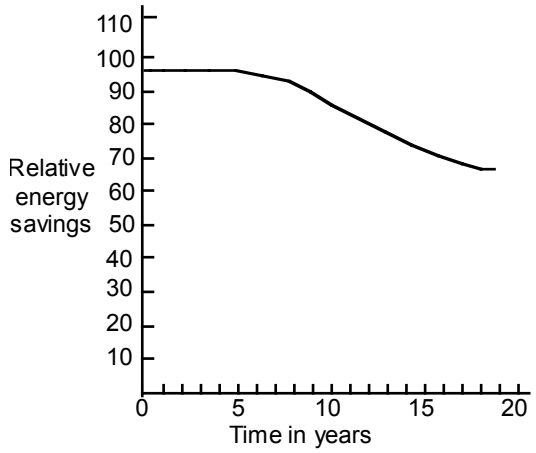


Figure 14. An Illustration of a Persistence Curve

Sustained behaviors are those that are repeated over and over – for example, an efficient motor installed as a result of program is replaced with an efficient motor when it dies.

An example of an organization using a technology to address the potential failure to sustain a behavior is the ENERGY STAR program. ENERGY STAR has been working with lighting manufacturers to develop a standard screw-in lamp socket adapter that accepts pin based CFLs. Once installed, the adaptor cannot be removed from the screw base and will only accept CFL pin base lamps. The idea is to commit the user to CFLs. From the perspective of energy efficiency, it is important to have a high degree of persistence and sustainability, at least until some newer technology or practice enters the market with greater benefits than the existing technology.

Although we have used examples from the business domain, the ideas of replication, emulation, spin-off, persistence, and sustainability apply across each of the domains. Attribution may reduce the amount of direct impact assigned to the program, but the program should also make sure that it counts impacts associated with replication, emulation, persistence, and sustainability. Impact evaluations need to examine each program-relevant domain to see if changes to the behaviors of participating actors result from program driven change, non-program driven change, or both, and also how actors' subsequent behaviors influence impacts. Not all relevant domains need to be examined at once.

6.3 The Double Counting Problem and Attribution

Under certain circumstances, there is a potential for double counting. For example, the California utilities operate appliance recycling programs. Some community-based programs supported by the utilities promote these program, as does the Statewide marketing campaign, Flex Your Power. Each of the programs is able to lay claim to some savings. The trick is for the evaluator to sort out the source of the savings (the attribution issue).

Community programs should be able to claim the incremental savings above what the regular utility program would generate in their community. Flex Your Power should be able to claim savings that result from its media efforts. The difficulty lies in determining what these savings may be. The problem is further complicated in one utility service territory where, without any promotion from the utility, some of the largest new appliance dealers use the incentives from the program to assist them in selling new refrigerators. They tell customers about the utility incentives and point out that they further reduce the cost of the transaction for the consumer.

Separating the effects entails understanding how each program attempts to influence the market and then determining from participants which of the sources influenced their decision. It is possible that the customer may have obtained information from multiple sources or may not know which sources they used. Getting the attribution correct is especially important when there are incentives

involved. For example, the utility, the contractor running the community program, and the companies running the media campaigns may all have performance contracts. A key question is whether the various programs are duplicating effort and claiming the same savings.

The software that DOE' ITDP develops provides another example of the attribution problem. Who can claim the savings: the software developer, the partners who run the training programs, the trainers who train the users, or the utility company that incentivizes company audits and whose contractors use the software? Each has a claim to some of the savings. One might argue that the software developer should be able to claim the incremental savings above standard practices that result from use of the software. After all, the software is convenient and credible and estimates and decisions get made that wouldn't otherwise get made. The sponsor should perhaps get credit for savings for people who get training who might not otherwise have learned about and used the software. The utility should get credit for audits and retrofits that might not otherwise have been done without the software. The problem isn't that each group shouldn't get credit; the challenge is that one needs to carefully attribute savings so that there is no double counting.

6.4 Evaluation Design Using Randomized Controlled Trials (RCTs)

Random controlled trials (RCTs) are the gold standard of impact evaluation design, but . . .

Attributing outcomes to a program is arguably the major challenge of an impact evaluation. The importance of this is demonstrated by OMB's guidance with respect to evaluation in the Performance Assessment Rating Tool (PART) (see the text box).

The "gold standard" for determining attribution is the *randomized controlled trial (RCT)*. In an RCT study design, people are randomly assigned to a participant (experimental) or control group. The subject does not know in which group he or she has been placed. This helps prevent the subject from acting according to his or expectations (placebo effect). Scientifically valid experiments treat control and experimental subjects identically and in isolation from one another.

Subjects in experiments are quite perceptive and may readily intuit in which experimental group they have been placed. Thus, in the medical world, where RCTs are most often used and the findings from such experiments can affect the health of millions of people, experimenters implement an additional double-blind procedure. The person or persons who deal directly with the subject do not know and are not told whether the subject is in the experimental or control group. This prevents the person working with the subject from providing subtle clues as to whether the subject is in the experimental or control groups.

OMB Evaluation Guidelines

Evaluations should be sufficiently rigorous to provide information on the effectiveness of the program. For programs that support or employ a range of services and approaches, evaluations should also provide information on the effectiveness of the various services and approach. To receive a *Yes*, agencies must demonstrate that they have chosen and applied evaluation methods that provide the most rigorous evidence of a program's effectiveness that is appropriate and feasible. A program may satisfy this criterion if the agency and OMB determine that the program is in the process of developing new evaluation approaches that will provide the most rigorous evidence possible by a specified future date.

The most significant aspect of program effectiveness is *impact* – the outcome of the program, which otherwise would not have occurred without the program intervention. A number of evaluation methodologies are available to measure the effectiveness of programs. Some, such as randomized controlled trials, are particularly well suited to measuring impacts. However, these studies are not suitable or feasible for every program, and a variety of evaluation methods may need to be considered. Other types of evaluations, including well-designed quasi-experimental studies, may provide useful information about the impact of a program and/or can help address *how* or *why* a program is effective (or ineffective) and should be considered when determining a yes or no answer. Quasi-experimental studies should be scrutinized given the increased possibility of an erroneous conclusion. (OMB, 2005, p. 28)

Figure 15 is a schematic of an RCT design. If we think in terms of an intervention for an energy efficiency program, the differences between before and after energy use, behavior, or attitudes of individuals in each group can be determined, and the results for the groups differenced to determine the net change or impact.

The reason an RCT is such a powerful design is that only the participant group is subject to the program intervention, while both groups are subject to any and all other influences equally. With random assignment, individual and external influences appear in both the treatment and control group. This design minimizes the potential for these influences to be an explanatory factor of study outcome. When you take the difference of the differences, you "net out" all other factors. This does not relieve one of having to statistically test to determine that the differences are significant.

However, RCTs can only be employed where one can control assignment to groups, keep the subjects unaware of their assignment, and isolate the subjects to prevent cross contamination. For example, it would be very difficult to apply RCTs to a Weatherization Assistance Program. In such a program, it is difficult to conceal the "treatment." Recipients are clearly aware that their homes have been weatherized and may modify behaviors to take advantage of the fact that they can change thermostat settings to make their homes more comfortable ("taking back" some of the savings). Further, ethical issues may arise as a result of not providing assistance equitably or not providing it to households most in need of weatherization. In the medical field, experimenters go to great lengths to address ethical issues, for example, working with specially selected populations and making sure that the patient has full knowledge of the risks and consequences.

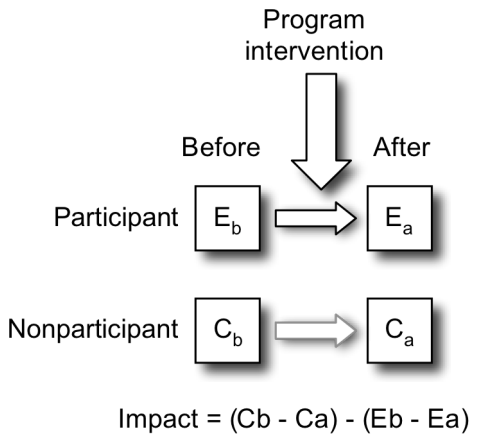


Figure 15. The Design of a Random Controlled Trial

RCTs are almost never used in energy research because of practical and ethical considerations.

At one point, there was substantial controversy because it appeared that the OMB was going to require that all evaluations of government programs be based on RCTs. The OMB now recognizes that RCTs are suitable or feasible for only some programs.

6.5 Comparison Group Designs — the Alternative to RCTs

The alternative to RCTs that is most commonly used in energy research is the comparison group design.

Evaluations of energy programs typically use alternative evaluation strategies. The most common alternative to RCTs is the *quasi-experimental method* that involves a comparison group rather than a control group. The subjects in a *comparison group* are not randomly assigned to a treatment condition, as are participants in a *control group*. In other words, the treatment group is self-selected by characteristics that differentiate them from others in a population.

An example of a quasi-experimental design is the evaluation of the FEMP, using a customer survey (Hall et. al., 2001a; 2001b; 2002). In one part of the study, Energy Service Performance Contracting (ESPC) participants and non-participants (people with similar job titles) eligible to participate in ESPC, were assigned to a stage of diffusion (see Chapters 2 and 3) before and after participation (or before and after an equivalent time period for non-participants) based on their responses to a series of questions in a survey. Table 4 shows the results. The percent of participants in the unaware, aware, and persuasion stages, declined by as much as 24 percent from before to after program participation (column 3). The percent of participants in the decision, implementation, and confirmation stages increased by an average of 16 percent. For non-participants, there was a substantial decline in those who were unaware (i.e., awareness increased), and modest increases among non-participants in the other stages. The bottom line is that participants moved rapidly in the direction of implementing and confirming behaviors following participation in the training program, while non-participants showed only small amounts of movement.

Table 4. Movement of FEMP ESPC Participants and Non-Participants through the Adoption Cycle

	Percent of participants (N=101)			Percent of aware non-participants (N=117)			True non-participants
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Diffusion stage	Before hearing about FEMP	Since involvement with FEMP	Change from before to after	Before hearing about FEMP	Since involve-ment with FEMP	Change from before to after	Percent of unaware non-participants
Unaware	24	0	-24	21	0	-21	63
Aware	27	10	-17	31	40	+9	24
Persuasion	12	7	-5	9	10	+1	5
Decision - no	10	7	-3	14	16	+2	1
Decision - yes	3	21	+18	8	9	+1	1
Implementation	7	24	+17	4	6	+2	4
Confirmation	18	32	+14	14	18	+4	2

Source: Hall, *et. al.*, 2002.

The big catch with comparison group designs is to identify the population from which the treatment group comes and then to draw a sample from it.

There are various strategies for selecting comparison groups. One method is to randomly select subjects for the comparison group from the same population as the treatment population. The catch phrase is "from the same population." Suppose that the goal is to create a comparison group for a weatherization assistance study. Applicants for weatherization assistance are from the low-income population, but they are not representative of that population. They represent a self-selected group since they volunteer to participate in the program, the characteristics of which may not be apparent. They are typically, but not always, homeowners. They may be having problems with their utility bills. They have the contacts, know how, and willingness to reach out, or they have contacts that help them make connections with a Community Action Agency delivering weatherization services. These characteristics only define some low-income households, so sampling from all low-income households is not sampling from an equivalent population.

An example of this problem comes from a study of payment behaviors for a Maryland electric utility bill payment program (Lee, et. al., 2007). The program was managed by the Maryland Public Utilities Commission (PUC) and administered by an agency inside the Department of Human Services. The sample of

participants was not a problem, because the study was ordered by the PUC. The difficulty was constructing the comparison group. An obvious possibility was to use households applying for other types of social services. However, because of confidentiality rules, information about persons not in the electric bill payment program could not be obtained. The alternative used was to randomly select households from the same nine digit zip codes from which the participants came. This resulted in a non-participant sample whose energy consumption was nearly identical to that of the participants but who were determined, on the basis of a survey, to be different in important characteristics such as home ownership, length of time in residence, etc. Although weighting was not used because of the small sample sizes, the households in the non-participant sample could have been weighted to more closely reflect the characteristics of the participant population to make the two groups more comparable.

Thus, it is difficult to define an equivalent population. Also, once the equivalent population has been defined, one has to obtain a list of members of the population to draw a sample. The zip code strategy presented a good strategy, because near neighbors could be identified on the basis of address and data anonymously obtained from billing records.

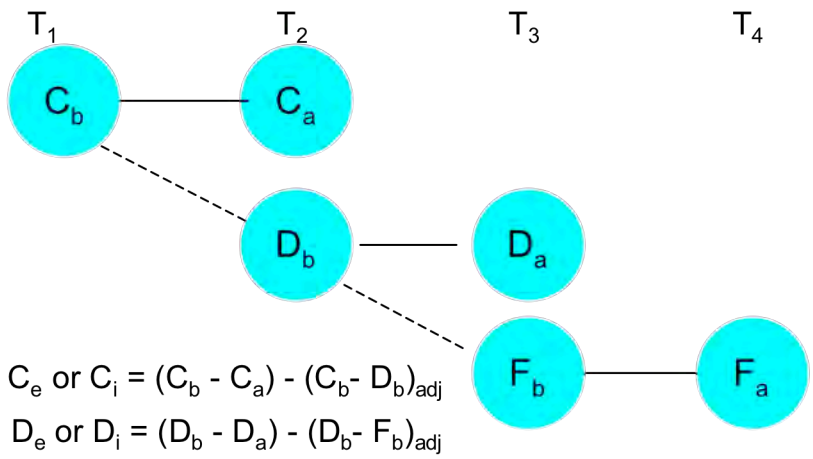
One way to skirt the equivalence problem is to draw samples of later participants and use them as a comparison group for earlier participants.

Some evaluations of weatherization programs have constructed comparison groups (called lagged comparison groups) by using later sign ups to form a comparison group for earlier sign ups (Figure 16) (Hall 1998; Reed 1999b; 1997). This should help with equivalence problems, but there is no guarantee that the characteristics of participants who signed up for the program two years previously are similar to households who sign up two years later. This might be particularly true if the program has matured and is more effective in reaching a target audience. The evaluator needs to determine the equivalence of the two groups on key characteristics.

Random selection of non-participants followed by careful matching on key variables is another solution to the comparison group problem.

There are many variations in developing a comparison sample. One approach is a sample matched on key characteristics. For the weatherization example, a group of low-income households might be randomly selected but then matched to

the characteristics of households in the treatment group by income, age, number of householders, structure of the household, historical energy use, etc. A frequent difficulty with this strategy is that there are not enough data to do the matching a priori.



T represents time
C, D, and F represent groups of firms
"b" and "a" represent before and after measurements
"e" and "I" represent market effects and impacts
Solid lines represent panels and dotted lines cross section comparisons

Figure 16. A Lagged Comparison Impact Design

The cases in a random sample of participants for a comparison group can be weighted on key characteristics to reflect the treatment group.

Another approach is weighting. The characteristics of the comparison group can be compared with the characteristics of the treatment group. Then, the distribution of individual characteristics of members of the comparison group can be weighted to match the distribution in the treatment group. To be done well, this requires substantial skill and some luck, as well as sufficiently sized samples.

Under certain circumstances, well matched comparison group studies can approach the rigor of RCTs. However, the use of comparison group studies does increase the risk of misleading results because of the difficulty in eliminating bias in the selection of the comparison group. For example, differences between the participant groups and comparison groups that affect their changes in energy use between the pre- and post-participation periods will bias the estimate of net

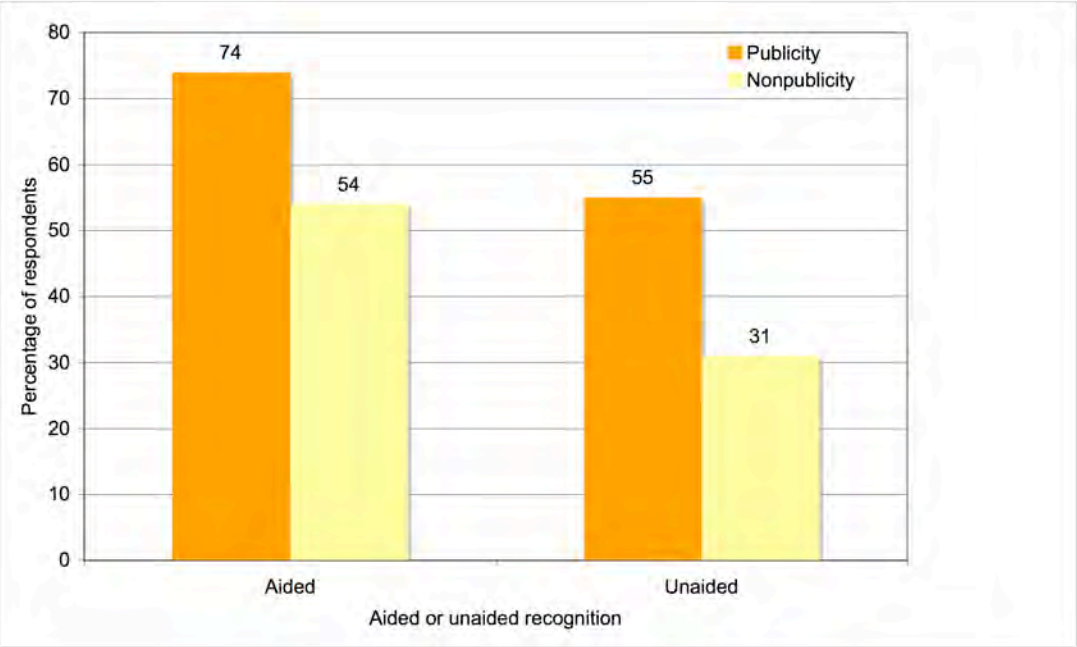
program savings (selection bias). Awareness of this risk is crucial to the design of such evaluations.

6.6 Other Comparison Strategies

One fairly easy way to compare participants and non-participants is to compare changes in measured energy use in different geographical areas where the target population and market characteristics are similar but the program application was different. For example, there may have been major technology demonstrations in one area and not in the other area.

Comparing two geographical areas, one that has been treated and the other which has not, is a way to establish a useful comparison.

Figure 17 is an example drawn from the annual ENERGY STAR Household Survey (CEE, 2005). To consider the effect of ENERGY STAR publicity on national awareness, evaluators implemented a survey and then compared *high publicity areas* (active local areas with ENERGY STAR programs recently sponsored by a utility, state agency, or other organization for two or more continuous years and sustained promotion and publicity from non-federal activities),



Source: CEE 2005

Figure 17. Aided and Unaided Recognition of the Energy Star Label in High Publicity and Low Publicity Areas

low publicity areas (Federal campaign activities only and no significant regional program), and *no publicity areas*. In the survey, respondents who said that they had heard of the ENERGY STAR label before the survey were asked to recognize it. *Aided recognition* was measured by showing the label and asking if the respondent had heard or seen it before. *Unaided recognition* was measured by asking this same question but without showing the label. Aided and unaided recognition was 20 percent or greater in the publicity areas compared to non-publicity areas. The conclusion is that publicity causes recognition of the Energy Star label, and the effect averages 20 percent.

If time series data are available, treatments and outcomes can be correlated – for example, media events and program sign-ups can be compared to see if the two types of events correspond.

Another method is to compare behaviors and program activity using time series data. If desired behaviors increase or decrease at the same time program activity increases or decreases, that is at least partial evidence that the program is working. For example, if the demand for refrigerator recycling increases and decreases when advertising increases and decreases, then that is an indicator that the program is reaching its intended audience and people are acting. If demand for refrigerator recycling is uncorrelated or partially correlated with advertising events, or is correlated with events other than advertising, this suggests that the advertising is not effective or perhaps partially effective.

Although rarely used in the energy field, one can create hypotheses and conduct experiments. For example, one might try four methods for recruiting customers to a program. One method might include an incentive and a brochure with a basic explanation of the program. A second method might not use incentives, but might provide a brochure asking people to adopt the technology or practice, appealing to the community good. A third might deliver the basic informational brochure. A fourth might offer an incentive and deliver a brochure that appeals to the community good. If the delivery method and the exposure to the brochure are constant and the experiment is implemented in similar populations, one should be able to infer information about the value of these various recruitment methods. Historic evaluations of demand-response programs suggest that the brochure with

the community appeal can be as effective as an incentive and basic information. Programs may want to do more with planned experiments especially when rolling out new programs, provided it is not difficult to isolate the target audiences.

It is strongly recommend that impact evaluations always use the program theory to predict the expected program outcomes. Furthermore, one goal of any impact evaluation should be to validate the program theory. With a weak comparison, or no comparison group, the evaluation can still make some assessment of the validity of the theory and the effects of the program. The program theory predicts a logical chain of sequential events that link program interventions to outcomes and impacts. If the evaluation is unable to demonstrate some (or all) of the linkages, this suggests that the particular linkage or the program did not work, or that the theory is wrong and needs to be revised.

In general, program managers and evaluators tend to find that their initial theories about how programs work are too simplistic and that the world of behavior is much more complex than expected. This is a positive, not a negative, finding. It means that when we do evaluations, we gain information that can help to produce a more effective program. Some of the most useful evaluations are those that show that a program has failed, because they reveal what should not be done and point to what might work. And, sometimes, programs do not work and should be terminated.

7. Design of Data Collection Activities

Data collection activities should be considered in the broadest context of program design. Because so many evaluations are designed after the fact, opportunities to collect data while a program is being implemented are often overlooked. Post-hoc collection of data is expensive, and participant recall or reconstruction of pre-existing data can reduce the accuracy of the information that is collected. Post-hoc collection of data also increases the cost of evaluations that, in turn, reduces the breadth and quality of the data that can be collected.

The following priorities are recommended as programs address their evaluation data needs:

- Developing, using, and improving routine data collection activities that will benefit the program and the evaluation.
- Making the best possible use of secondary data sources.
- Engaging in primary data collection.

Recommended priorities are improving and using routine data collection, making use of secondary data sources whenever possible, and then undertaking primary data collection.

7.1 Routine Data Collection

Many programs routinely collect large amounts of data. Some programs already collect data that could be used for evaluation purposes. The ITDP's IAC is a good example. This center maintains audit information from industrial plants in a database. There is follow-up with firms that have audits to find out which recommended measures are implemented. This information can be used to estimate gross savings, and, with a lagged comparison group design, might also provide estimates of net savings. It would, of course, be useful if the audit savings estimates could be verified to see what the realization rate is (i.e., actual savings divided by estimated savings).

Programs routinely collect a lot of data. Improving the way in which that data are collected will make it useful for evaluation and for other purposes, such as marketing.

For example, a study of record keeping practices of selected EERE deployment programs resulted in Table 5, which shows the distribution of the types of records being kept across programs (Schweitzer and Saulsbury, 2005). Although program records tend to be largely comprised of output data that track various

program activity levels, some data that would support analysis of outcomes are collected from program participants. Table 6 indicates how various types of deployment program output and outcome records were categorized according to the five stages of the diffusion of innovations (awareness, persuasion, decision making, implementation, and confirmation). Program output records covered all five stages. However, program outcome records were limited to implementation and confirmation stages.

Table 5. Types of Program Records Most Commonly Collected

Types of Records	Number of Programs[9]
Inputs	
Program funding by activity/project	6
Funding provided by program partners (e.g., state and local government grants, private sector funding)	4
Total program funding	3
Project/activity budget and milestones	3
Organizational information about program partners (e.g., name, address, contact, organizational goals)	3
Number/amount of federal grants distributed to program partners (e.g., state governments, local governments, utilities)	2
Outputs	
Outreach/educational events conducted by the program and/or program partners (e.g., workshops, seminars, meetings, conferences)	5
Technical assistance provided by the program and/or program partners	5
Standard reports on program activities, events, and results (e.g., monthly reports, quarterly reports, semi-annual reports, annual reports, JOULE reports)	5
Outreach/educational materials produced and distributed by the program and/or program partners (e.g., publications, software, videos)	4
Training sessions conducted by the program and/or program partners (e.g., number of sessions, dates, locations, topics, attendance)	4

[9] A total of 11 programs were reviewed.

Types of Records	Number of Programs[9]
Reports on training (for DOE use)	2
Reports on case studies and success stories	2
Outcomes	
Energy savings (in Btu, kWh, petroleum displacement)*	9
Cost savings (in $)	7
Short term program outcomes (e.g., number and type of units installed/weatherized/in inventory)	4
Installed capacity	2
Cost effectiveness of recommended/implemented measures	2
Reduction in air pollution/CO2 emissions	2
Reduction in waste due to increased process efficiency	2
Resources leveraged (e.g., funding, in-kind support)	2

Reference: Schweitzer and Saulsbury, 2005

*Energy savings are based on self reports and estimates.

Table 6. Program Output and Outcome Records by Stages of Diffusion

Types of Records	Stage(s) of Diffusion
Outputs	
Outreach/educational events conducted	Awareness; persuasion or information seeking; decision making
Technical assistance provided	Persuasion or information seeking; implementation
Standard reports on program activities, events, and results	Decision making; confirmation
Outreach/educational materials produced and distributed	Awareness; persuasion or information seeking; decision making
Training sessions conducted	Decision making; implementation
Outcomes	
Short term program outcomes (e.g., number and type of units installed/weatherized/in inventory)	Implementation
Energy savings*	Implementation; confirmation
Cost savings	Implementation; confirmation
Short term program outcomes	Implementation

Types of Records	Stage(s) of Diffusion
Installed capacity	Implementation
Cost-effectiveness of recommended/implemented measures	Decision making; implementation; confirmation
Reduction in air pollution/CO2 emissions	Implementation; confirmation
Reduction in waste due to increased process efficiency	Implementation; confirmation
Resources leveraged	Confirmation

Reference: Modified from Schweitzer and Saulsbury, 2005

*Energy savings are based on self reports and estimates.

7.2 An Example of Implementing Routine Data Collection

Routine data activities can often be enhanced to produce the types of data needed for evaluation. The following is an example of how that might be done for training. The hypothetical example is based on training for the ITDP's Steam System Design Tool, but it is appropriate to almost any training program.

For evaluation of a training program, augment the usual questionnaire by asking what the trainee's practices are before they take the training. This provides a baseline and will help the instructor to know how to present the course.

Like other EERE programs, the ITDP conducts many training events that include an evaluation at the end. Currently, the evaluations tend to focus on the quality of the instruction and the comfort of the training facility. This is useful information, but the questions that need to be answered are: (1) how the knowledge gained through the training gets used, if at all; (2) whether that knowledge leads to changes in facilities and the operation of facilities; and (3) how those changes effect energy consumption and emissions.

Applying principles of good design, the evaluation of training plan should:

1. Focus on behavior.
2. Make use of the training as an opportunity to gather data.
3. Collect behavioral data from before and after training.
4. Include some kind of comparison group.

To assess market effects and energy saving impacts of steam system training, we need to (1) identify behaviors and system configurations that represent efficient operation of boilers; (2) determine if the relevant configurations and behaviors were in place or practiced before and after the training; (3) identify configura-

tions and behaviors that changed; and (4) estimate the savings from the changed configurations or behavior.

The basic principles of efficient steam systems are easily extracted from ITDP's documents and software. Examples of selected steam system configurations and relevant behaviors consistent with efficient operation are recorded in the material:

- Boiler efficiency is determined.
- Hot spots on the shell are identified.
- Flue gas oxygen content is monitored.
- Flue gas oxygen content is maintained within limits.
- Condenser heat transfer components are clean.
- A schematic of the system is available.
- A system mass and energy balance is available.
- Steam flow through vents and pressure reducing stations is understood.
- The effects of changing boiler pressure are understood.
- Piping is monitored for steam leaks.
- A steam trap database exists.
- The operation of steam traps is checked annually.
- Traps are of proper size and type.
- Condensate loops are clear.
- Pipes are insulated.

After the training, conduct another survey to find out how practices have changed and whether they are attributable to the training course.

To evaluate market effects (changes in practices and equipment), energy savings, and energy intensity, the basic steps in an evaluation should proceed as follows:

1. Develop a behavioral questionnaire focusing on the identified behaviors.
2. Administer the questionnaire prior to the training. The questionnaire could be sent to attendees or administered at the very beginning of training. The questionnaire assists the trainer in understanding the information needs of the trainees and serves as a baseline for the evaluation.
3. Administer an augmented questionnaire to trainees between six and eighteen months after training.[10] Repeat the same questions as in the original

[10] If changes to the system involve capital, then time has to be allowed for the acquisition of capital in addition to the time it takes to implement the changes.

survey but also ask about specific modifications to the system and operation of the system that have been made since training. Also, ask for estimates of savings and production rates, so that energy intensity can be measured. Include some questions to assess whether the changes were made in response to training and whether the changes would have been made in the absence of the program. Determine the number of boilers for which systems changes were made at this and at other sites.

Attribution needs to be evaluated. For example, with rising fuel prices, program participants might pursue improvements to boiler systems anyway. Perhaps the firms use consultants who would have recommended changes without the training. From an attribution standpoint, the program can take credit for changes that were directly a result of its activities or a derivative of those activities or changes that occurred earlier than they would have.

Find a way to create a comparison group.

Using the lagged comparison evaluation design outlined in Section 7.4, the baseline data for a group of new participants who receive training at about the same time as the post-training survey is conducted can be used as a proxy for a post-comparison group measurement. In this scenario, the baseline participation of the original group is compared to the baseline for the subsequent group of participants to establish a proxy for what changes might have occurred without the program. This has the benefit of minimizing data collection costs, and it samples from a group of non-participants who will use the program.

In this instance, you have a panel design (before and after measurements for the treatment group) and a cross-sectional design for the comparison group. As noted above, this is less than the ideal, but it should provide some insight into changes induced by external trends (such as energy prices that might influence outcomes). It is less than ideal because new participants who take the training 18 months after the original group may differ in significant ways from earlier participants. By doing matching and weighting, it may be possible to construct a synthetic comparison group that is very much like the original group of participants.

Following this design, the data collection and analysis will result in a table that looks something like Table 7. In this illustration, training participants, when

compared to non-participants, were less likely to have engaged in desirable behaviors prior to the training, but subsequent to the training their desirable behaviors exceeded those of the non-participants. The net difference in behavior is 22 percent for maintaining flue gas oxygen content within limits, and 15 percent and for cleaning steam traps annually. These data demonstrate that the behaviors of participants improved. With additional information about the facilities, these improvements could be converted into estimates of energy savings.

Table 7. Hypothetical Measurements of the Percent of Participating and Non-Participating Facilities Engaging in Two Critical Steam Systems Behaviors Before and After Training

Practice / Measure		Before	After	Difference	N
Flue gas oxygen content is maintained within limits.	Participant	47	74	27	91
	Non-participant	61	65	5	52
	Net difference	-14	9	22	
Steam traps are cleaned annually.	Participant	18	38	20	91
	Non-participant	26	31	5	52
	Net difference	-8	7	15	

The point of this illustration is to suggest that program managers can do much more with routine data collection than they may be doing. In this case, the changes to routine data collection were to modify the content of the existing training evaluation instrument, add a post-participation survey that covers the same questions and obtains information about what might have been implemented, and make sure that the initial training questionnaire contains enough information so that subsequent trainees can be used as comparison groups. Potentially, this is an effective but low budget strategy. At some point, the evaluator may want to undertake an analysis to see if subsequent groups of trainees are an adequate comparison group.

7.3 Secondary Data Sources

Secondary data sources (Table 8) are also available for use in program evaluation. These types of data may be particularly useful for market studies but may also be of value in impact evaluations. Each data source has its strengths and weaknesses and is suited for different purposes.

Table 8. Examples of Secondary Data Sources

Types of Records
Websites/Trade Publications
Top 25 building owners
Top 25 building management firms
Top 100 school districts
Operation of retail food sales
Public Data Sets
Commercial Building Energy Consumption Survey (CBECS)
Residential Building Energy Consumption Survey (RBECS)
State Licensing Board data
US Census (Housing, Population)
Manufacturing plant data
Commercial Data Sets
Dun and Bradstreet (companies)
F.W. Dodge (construction)
Harris (industrial firms)
PRIZM (market segments)

Source: Reed 2004b

The Commercial Building Energy Consumption Survey (CBECS) and the Residential Building Energy Consumption Survey (RBECS) samples mentioned in Table 8 contain large numbers of buildings. Some of the data in CBECS and RBECS are obtained through surveys, some from on-site visits, and some from utility records. Thus, there are limitations with these data, especially with respect to energy use. Also, these are national samples of buildings, making it difficult, if not impossible, to develop location-specific estimates of energy uses.

California recently released the data for its Commercial End-Use Sample (CEUS) (Itron 2006 http://www.energy.ca.gov/ceus/index.html). The data for this sample were collected on site. An audit was completed for each structure. Utility data are available and the energy use of each structure was modeled. These will be extremely valuable, although these data will be limited as well, because of the cultural and climate constraints imposed by a California-based sample.

7.3.1 Primary data collection

A variety of methodological options exist for collecting data. Table 8 in the *EERE Guide for Managing General Program Evaluation Studies* lists the more common methods, along with several of their key characteristics. For ultimate outcomes, they include on-site metering and other types of measuring equipment, building simulation modeling, and utility billing data. These data collection techniques are also often accompanied by in-person and focus group interviews and surveys: in-person (onsite visits), mail, and telephone. These techniques can also be used for examining market effects. Another source of data for assessing market effects is program records and reports. Some examples of primary data sources are provided below:

1. Annual surveys that track awareness, understanding, and influence on purchases (e.g., ENERGY STAR household survey)
2. Regional market characterization studies that periodically track stock, purchase and sales practices and decision making (e.g., retailer interviews, out-of-state retailer interviews, mystery shopper store surveys, floor stock store surveys, general population surveys, purchaser surveys, out-of-state purchaser surveys, and program adopter surveys)
3. Sales and market share tracking and cross-sectional analysis (e.g., the Association of Home Appliance Manufacturers tracks shipment data at the county level)

OMB clearance is needed if data are to be collected from outside the Federal government from more than nine individuals (see the text box). When data are collected by use of surveys or protocols, the program implementer or the evaluator will need to prepare an Information Collection Request (ICR) and formally submit it to OMB for clearance.

The delay can be problematic. It is recommended that programs build a strategic evaluation plan. Based on the plan, programs can develop sets of generic questions that can then be used in the construction of specific questionnaires.

It may be possible to get clearance for a generic set of questions that can then be repeated in more specific questionnaires. By referencing a strategic plan, it may be possible to clearly identify the need for and purpose of generic questions.

OMB Clearance to Collect Data

As noted in the Evaluation Guide, OMB clearance is needed if data are to be collected from outside the Federal Government.[11] Specifically, if the audience from which you need to collect data does not consist exclusively of Federal Government employees and the evaluation needs primary data from more than nine members of this audience, including potential customers, then the data collection activity will receive clearance from the OMB. Federal Government employees are excluded from the OMB clearance requirement only if the questions to be asked of them involve activities associated with their employment; otherwise, surveys of Federal employees also require OMB clearance (*EERE Guide*).

As noted in the Evaluation Guide, the time required to obtain OMB clearance varies and may hinder the timely performance and use of program evaluation:

- **For customer satisfaction surveys and pretests of other survey instruments,** there is an expedited process that, in most cases, takes two to four weeks.[12] The Forms Clearance staff of EIA's Statistics and Methods Group can assist DOE staff with this process.
- **For surveys other than customer satisfaction surveys,** the OMB clearance process takes longer. Currently, the entire clearance process may require five to eight months. Clearance applications are submitted to the Records Management Office (IM-11) of DOE's Chief Information Officer.

An OMB clearance is valid for three years.[13] Additional information about how to obtain an OMB clearance for data collection is found in Appendix A6 of the Evaluation Guide.

[11] The legal requirement for OMB clearance of a survey is the Paperwork Reduction Act of 1980 (amended 1995). OMB requires that its clearance review take no less than 30 days and no more than 60 days; however, the entire process can take five to 12 months to receive an approval. An expedited process exists for customer satisfaction research.

Barring a generic clearance, it may be less difficult to get questions cleared that have previously been cleared and used. Finally, the strategic plan and the generic questions also make it possible to anticipate the need for clearance and to get clearance procedures started in a timely fashion.

[12] OMB approved generic clearance DOE-887.
[13] Clearances resulting from emergency reviews last six months; however, emergency reviews are unlikely to apply for general evaluations.

8. Reporting Outputs

A variety of evaluation reporting outputs can be generated with the approach described in Chapters 2 through 7. A number of illustrative examples are provided below.

8.1 Energy Savings Reporting

An example of an energy savings report from assessments of industrial plants is shown in Table 9. The ITDP completed plant-wide assessments (PWAs) at eight facilities. These assessments involved sending personnel on site to assist facility managers and consultants to identify changes to technologies and practices that would increase energy efficiency. Subsequent to the on-site visit, the firm was responsible for selecting and implementing the recommendations.

The ITDP used consultants to follow up with the firms to determine what recommendations had been implemented and whether similar activities had been undertaken at the same or another plant owned by the firm (replication). The recommendations ranged from simple changes to operational practices that could be implemented almost immediately to recommendations that would require capital investment. For changes requiring capital investments, the firms typically developed internal proposals that competed for funding with proposals for other initiatives at the plant. The recommendations that were implemented for eight plants resulted in total savings of 5,482 BBtu/year.[14]

Table 9. Energy Savings (BBtu/yr) from Plant-Wide Assessments (PWA) and Replications, FY03

Total original PWAs	8
Savings generated for original PWAs (BBtu/yr)	5,482
Total replicated PWAs	22
Savings generated for replicated PWAs (BBtu/yr)	5,606
Total energy savings for PWA and replications (BBtu/yr)	11,088

Source: Martin, 2004.

[14] The number 5,482 seems quite precise. Given the methods for calculating the savings 5,400 or even 5,000 BBtu/yr might be adequate and signal the level of precision of the estimate.

As noted above, an important goal of EERE deployment programs is to encourage replication, emulation, spin-offs, and sustained behavior. This study clearly illustrates the importance of replication effects. After completing the initial project, these firms completed 22 projects that were similar to the one that was the subject of the PWA. These projects could have been in the same or different plants. These additional 22 projects more than doubled the savings from the original audits. This study did not control for external factors, so the net savings may be less. Nonetheless, the program clearly triggered additional projects.

A second example of an energy savings reporting output is to estimate and report results by both program activity and delivery channel. For example, Table 10 shows that energy savings are expected to be greater with direct project assistance, compared to interventions based on software and publications, for energy projects in industrial facilities.

Table 10. Illustrative Example of Energy Savings (kWh/year) for Key Program Activities and Delivery Channels

	Project Activities								
	Design Assistance	Audits	Industrial Facili-ties	Sustainable Facilities	Combined Heat and Power	Operations and Maintenance	Renewables	Technical Assis-tance	New Technology Demonstrations
Direct Project Assistance	1,000	300	800	200	200	100	200	200	200
Energy Assessments	500	200	500	100	100	75	100	100	100
Training and Workshops	100	75	200	50	50	20	50	50	50
Software and Publications	50	25	25	25	25	5	25	25	25

8.2 Market Effects Reporting

The reports from evaluations of market effects will vary, depending on the type of approach used for assessing the diffusion of innovations.

8.2.1 Example for the product characteristics model

The importance of the product characteristics model (Section 2.6) is illustrated in Table 11. In-depth interviews with 10 of the 56 largest retail food sales firms in the United States revealed important linkages between the characteristics of the firm and characteristics of new and/or innovative energy efficiency technologies that could be installed in supermarkets (Reed, Johnson, *et.al.*, forthcoming 2007; Johnson, 2007; Reed, 2006a).

Table 11. Energy Efficiency Product Characteristics, Product Criterion, and Importance for Three Types of Firms

		Three types of firms		
Energy efficiency product characteristics	Criterion	Merchandising oriented, large number of widely separated stores	Merchandising oriented, modest number of geographically proximate stores, internal maintenance staff	High volume, operationally oriented, large number of widely separated stores, contracted maintenance
Relative advantage	Good payback	X	X	X
	Low maintenance	X	X	X
Compatibility	High image / high aesthetic	X	X	
	Low customer barriers	X	X	
Complexity	Willingness to manage high complexity		X	
Trialabiltiy	Trialability	X	X	X
Observability	Observability		X	

Source: Reed, Johnson, *et. al.*, forthcoming 2007.

The reader may recall that there are five product characteristics: relative advantage, compatibility, complexity, trialability, and observability. In terms of relative advantage, these grocery firms focused on good payback and low maintenance. These firms tend to be oriented to one of two strategies, merchandizing (i.e., creating a positive customer experience) or operational execution (i.e., moving high volumes of product through the doors at low cost). Innovations must be compatible with a company's overall strategy in the market. The managers who were interviewed were concerned about the complexity of installing and operating equipment and the requirements that complexity might impose on them. Finally, all of these firms required pilots (trialability) and all were monitoring the performance of energy use in stores at some level (observability). Some were much more heavily involved with monitoring store systems than others.

Table 11 shows how product characteristics vary for three supermarket types. These types are generic, although we can categorize specific supermarket chains in relation to each of them. All three types agree that the desirable paybacks for equipment is three years or less, preferably two. There are some exceptions on this score among respondents, but the exceptions are not relevant here. They all agree they are looking for low maintenance technologies. These two criteria can be traded off. For example, several of the respondents in the study observed that they might be likely to install refrigerator case LED lighting in the future, but they would do so for maintenance reasons rather than energy savings reasons. In addition, representatives agreed that equipment must be proven through pilot tests before it can be included in store prototypes.

An example of the first type of store is a large regional chain with more than 1,000 stores. Merchandising is important for this store. Thus, the merchandising department can veto efficiency measures that are cost effective if it perceives that the changes might make the store less attractive to customers (i.e., there is concern that energy-efficient lighting may not be able to highlight the colors in packaging). Also, this store does not have windows between the meat preparation area and the customer area, because merchandising wants meat cutters to be accessible to customers. Since the meat preparation area is kept at 47°F and the

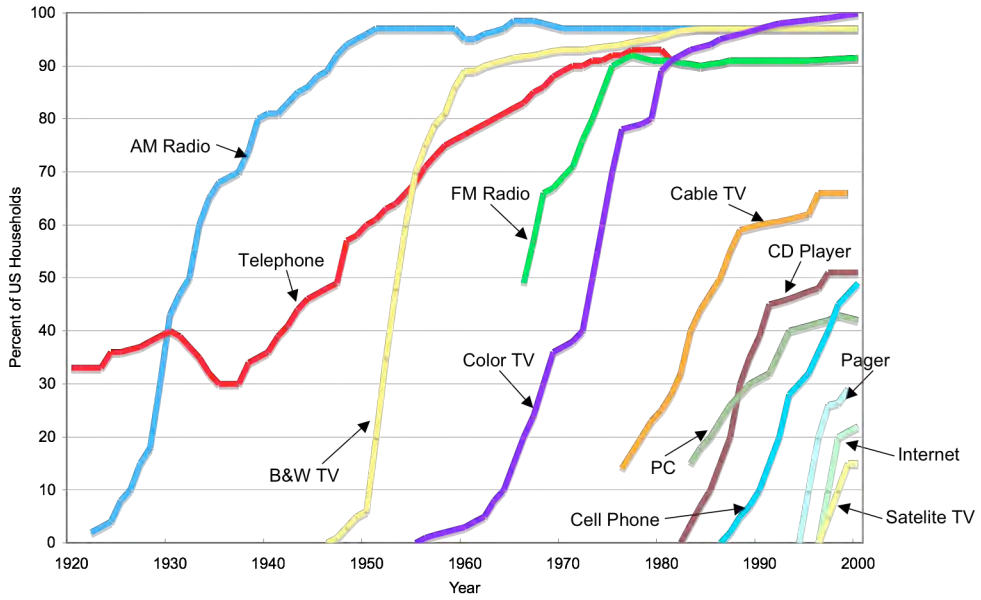
customer area in the 70s, this results in increased energy consumption. Store three with its volume and operational orientation is much less concerned with the aesthetics of lighting and, therefore, uses energy efficient high bay lighting (which doesn't necessarily display all goods to advantage) and has windows that separate meat preparers from the main part of the store.

The second store is a smaller regional entity with more than 100 stores with a fairly compact market area. Like the first store type, merchandising is important for this firm, and changes must meet with the approval of the merchandising department. This firm is different than the other two in its tolerance for complexity. This firm is doing a great deal with energy efficiency, especially in the area of refrigeration. It is varying floating head pressures across a much wider temperature range than do other firms, and it has done a significant amount on its own to integrate systems – e.g., doing heat recovery from the refrigeration systems to heat the building and provide hot water for wash wares. This firm also has an extensive central monitoring system that allows it to run its systems at finer tolerances. The facility manager said that he is able to do this because he has trained technicians who understand how to maintain and manage the equipment. The other firms contract out most of their maintenance.

How does this relate to product characteristics? Larger firms are more likely to want turnkey systems with minimal complexity. Also, complexity needs to be engineered out of monitoring and control systems if integrated systems are to be widely adopted.

8.2.2 Example for the adopter characteristics model

The number of people adopting an innovation (e.g., electronic ballasts) is approximately normally distributed around the average length of time for people to adopt a particular innovation (a bell curve, see Figure 4). If the number or percentage of adoptions is accumulated through time, rather than reported as a distribution, the result is an S-curve. Figure 18 shows S-Curves for a variety of technologies over the 80-year period between 1920 and 2000 (Fallah, 2006).



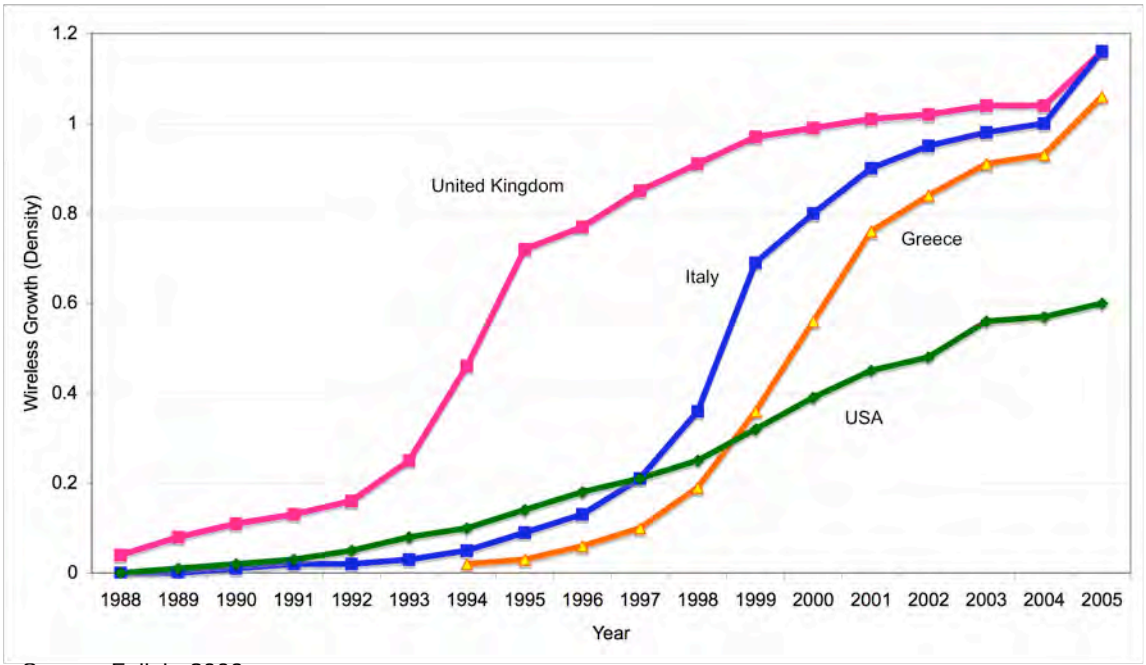
Source: Fallah, 2006

Figure 18. Penetration of Selected Technologies in the United States between 1920 and 2000

Comparisons among the curves are interesting. Radio took off and continued a steep rise even during the depression years of the 1930s. On the other hand, the penetration of telephones declined during the depression possibly indicating that is was considered a luxury. There was an 85 percent increase in the penetration of black and white televisions in five years. It took nearly 15 years for color television to reach the same level of saturation. Getting a television was probably more compelling than getting a color television. Although not shown in this graphic, the percentage of black and white televisions has declined substantially (Fallah, 2006).

Figure 19 shows S-curves for wireless telephones (telephones per person) for three European countries and the United States. Notice that the United Kingdom led the other three countries. In the United Kingdom, wireless technology began to take off about 1992 and reached more than half of the population by 1995. The

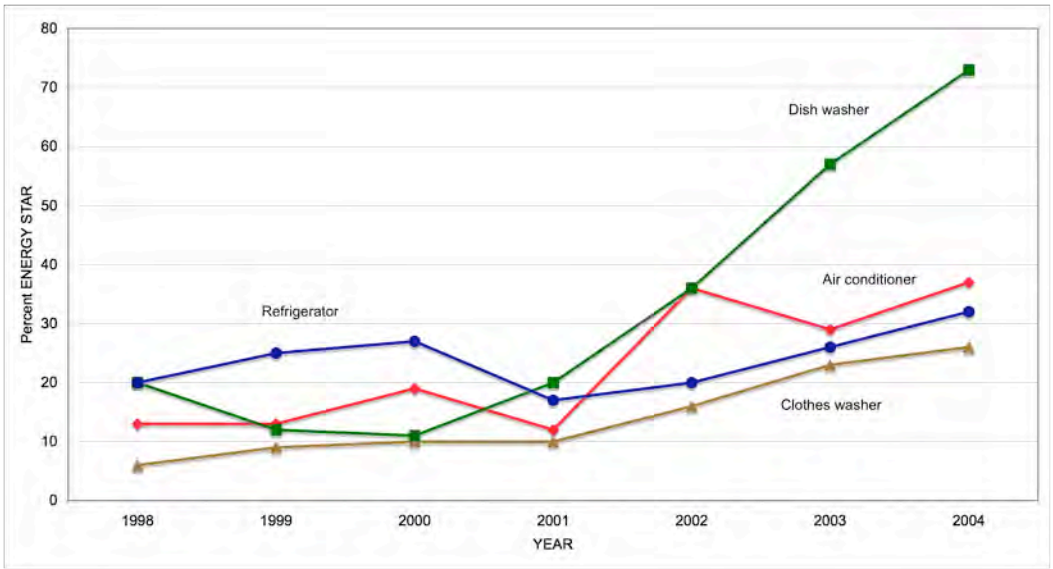
curves for Italy and Greece are quite similar to each other. The takeoff in Italy started about 1996, and in Greece in 1997. The similarity is attributed to similarities in the market environment and demographics of the population. The growth in wireless technology has been much slower in the United States probably because of the existing landline infrastructure and the more dispersed population affecting the economics of delivery (Fallah, 2006).



Source: Fallah, 2006

Figure 19. S-curves for Wireless Telephone Technology

Figure 20 illustrates the penetration of ENERGY Star products. ENERGY STAR dishwashers have clearly taken off and represented 72 percent of the market by 2004. Clothes washers, refrigerators, and air conditioners exhibit a much less dramatic trend. The percent of the market captured by Energy Star refrigerators has tended to oscillate some with a decline in purchases in 2000 and 2001. The slower penetration for refrigerators may be attributable to the appeal of other desired product features that override consideration of the approximately $20 annual savings from running an ENERGY STAR as opposed to a non-ENERGY STAR refrigerator.



Source: Fallah, 2006

Figure 20. Market Penetration of ENERGY STAR Products 1998-2004

8.2.3 Diffusion process example

An example of the results of using the diffusion process model is shown in Figure 21. This figure is based on an evaluation of FEMP's Technical Assistance Program (Hall et al., 2002). It shows the position of FEMP customers in the adoption cycle before and after receiving specific FEMP training. The left side indicates the customers' position in the adoption cycle prior to participation. The right side indicates their position

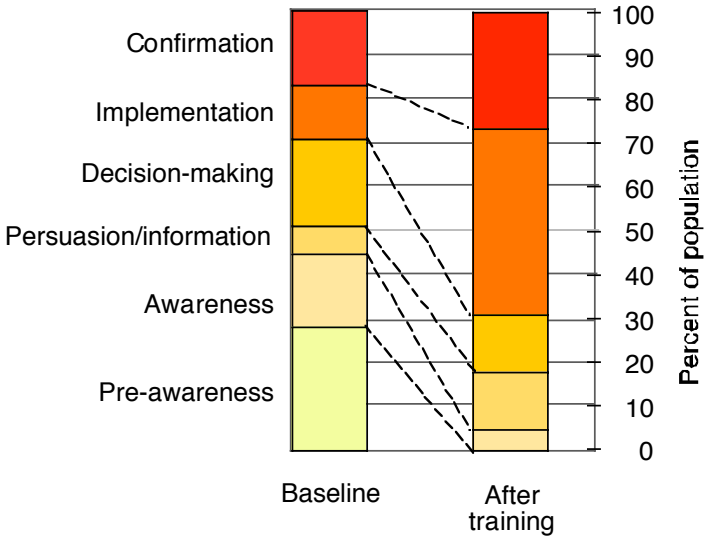


Figure 21. Movement of FEMP Customers through Stages of Diffusion as a Result of Training (Hall, 2001)

approximately one year after participation (training). As indicated in the figure, 29 percent of FEMP customers were unaware of the recommended energy efficient, water efficient, or environmentally friendly product, technology, or concept prior to exposure from FEMP. Sixteen percent had just become aware of the technology before FEMP exposure, and eight percent had already begun collecting information about the technology. Nineteen percent were in the process of making a decision about using the technology at the time of contact, and 12 percent had already implemented the FEMP recommended technology in their organizations. An additional 16 percent were in the confirmation stage and had repeatedly implemented the technology prior to the FEMP assistance targeted in the survey.

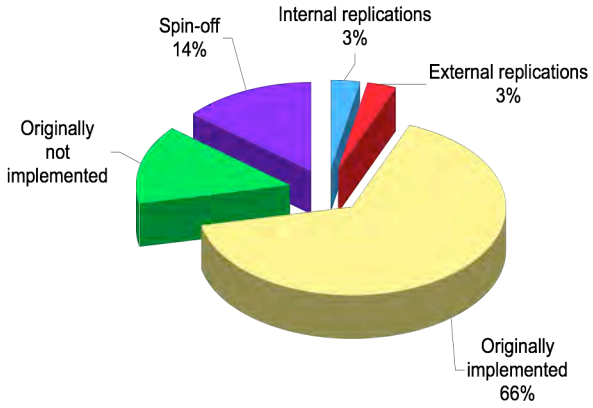
Following FEMP assistance, no customers remained unaware of the recommended technologies as a result of the FEMP programs, and only six percent indicated that they were just becoming aware of the technology. A similar type of movement occurred at the top end of the diffusion scale. Prior to FEMP assistance, 28 percent of all customers were in the implementation or confirmation stage. This increased to 67 percent following FEMP assistance.

8.2.4 Confirming behavior examples

Two examples of confirming behaviors are shown in Figure 22 and Figure 23. Figure 22 is based on an evaluation of the EERE's Industrial Assessment Center (IAC) Program (Martin et al., 2003). The IAC Program provides funding to operate 26 assessment centers across the United States, and these centers assess energy use, waste, and production activities at client manufacturing plants and generate recommendations for saving energy and reducing operating costs. The IAC Program collects information on IAC assessments and individual assessment recommendations. In this evaluation, the recommendations were classified into five types: originally implemented, originally unimplemented, unimplemented, internal replication, external replication, and spin-off. The designations "originally implemented" and "originally unimplemented" are recorded in the IAC Program database on the basis of a client's indication of whether they intended to

implement recommendations. IAC clients sometimes also implement recommendations either at another location in the assessed plant (*internal replication*) or at another plant related to the original plant (*external replication*). Clients also sometimes implement *spin-off measures* – measures related to but different from the original recommendations.

Figure 22 shows the composition of source energy savings by the five savings types. Although most of the savings are from originally implemented recommendations, substantial contributions are also provided from spin-offs, replications, and implementations of originally unimplemented recommendations.[15]



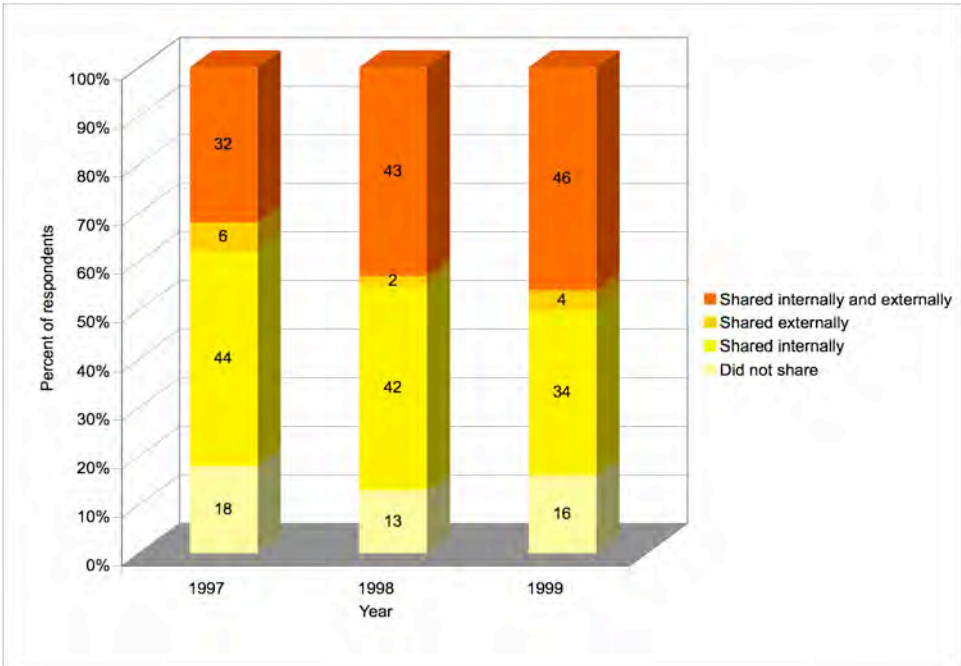
Source: Martin 2003

Figure 22. Implementation, Replication, Spin-off, and Delayed Implementation in the IAC Program

Figure 23 is based on an evaluation of FEMP's Technical Assistance Program (Hall et al., 2002). Sharing FEMP information with others has the potential to multiply FEMP's impacts within their target markets and greatly speed market transformation. Sharing information (or customer networking) is identified as the single most effective method for speeding the diffusion of innovations (Rogers, 2003). If FEMP customers share their information with other professionals inside

[15] Sometimes clients change their minds and decide to implement recommendations that were originally rejected.

or outside of their organizations, FEMP's impacts can be multiplied across markets and within organizations. As shown in this figure, there was an increase in the percent of participants sharing FEMP technical assistance information both internally (within their organizations) and externally (outside of their organizations). The percent sharing both internally and externally increased from 31 percent in 1997 to 43 percent in 1998 to 46 percent in 1999.



Source: Hall, 2002

Figure 23. Confirming Behaviors through Participant Information Sharing

8.2.5 Sustaining behavior example

One of the goals of deployment programs is to change behaviors so that they are sustained over a long period of time. Table 12 displays self-reported responses to the Energy Star label in four years (Cadmus et. al., 2003; Cadmus and KEMA, 2004; Cadmus and KEMA, 2005; Cadmus and KEMA, 2006). The first row shows the trend for purchasers of Energy Star appliances who were able to recognize the Energy Star label after some prompting and who said that they were influenced to purchase the appliance by the label. The second row describes trends in the recognition of the Energy Star label without assistance. The primary

point of this table is that levels of influence of the label on respondents' purchase decisions and recognition of the meaning of the label increased over time. These data suggest that the Energy Star program's efforts to increase brand awareness and influence are showing success.

Table 12. Influence of the Energy Star Label (Information/Persuasion) on Purchase Decisions

Energy Star Nationwide Results	2002* percent	2003 percent	2004 percent	2005 percent
Respondents who said that they had been somewhat or very much influenced by the label, who had an aided recognition of the label, and were purchasers of Energy Star appliances	4	52	54	63
Respondents who exhibited a general to high-level of understanding of the meaning of the label when the label was recognized unaided, or if not recognized unaided, the respondent was able to intuit the meaning of the label upon seeing it	58	63	68	70

Source: Cadmus et. al., 2003; Cadmus and KEMA, 2004; Cadmus and KEMA, 2005; Cadmus and KEMA, 2006

- The sampling procedure in 2002 was different than that in 2003 and 2004, so the results are not fully comparable.

8.2.6 Network example

Previously, we have referred to one-to-one communication and network analysis. Figure 24 shows a simplified network of interconnections among architectural, engineering, and retail planning firms for "The Limited" retail store (Reed, 2003a; 2003b; 2003d). Using a network analysis tool and the FW Dodge Players data for New Jersey enabled us to establish the linkages among firms. There are three important points to be drawn from this analysis. The first is that the firms represented by red nodes are all located in or around Columbus, Ohio, where The Limited headquarters is located. Although these data were taken from New Jersey, the red nodes represent the firms The Limited uses to do the centralized design of its stores. The green nodes represent firms that do construction in

New Jersey. The unfilled red circle represents HBK, a national engineering firm. HBK is also linked to the Gap, another large retail firm. The key point to be drawn from this diagram is that to influence the design (or energy efficiency) of The Limited stores nationwide, one needs to focus on a few firms in Columbus. Secondly, the HBK connection represents a path for cross fertilization. To the extent that engineering firms can carry information about energy efficiency among firms, HBK can be a carrier. The extent to which that can happen may be limited by whether or not information is considered proprietary, and the extent the various offices of HBK exchange information among themselves.

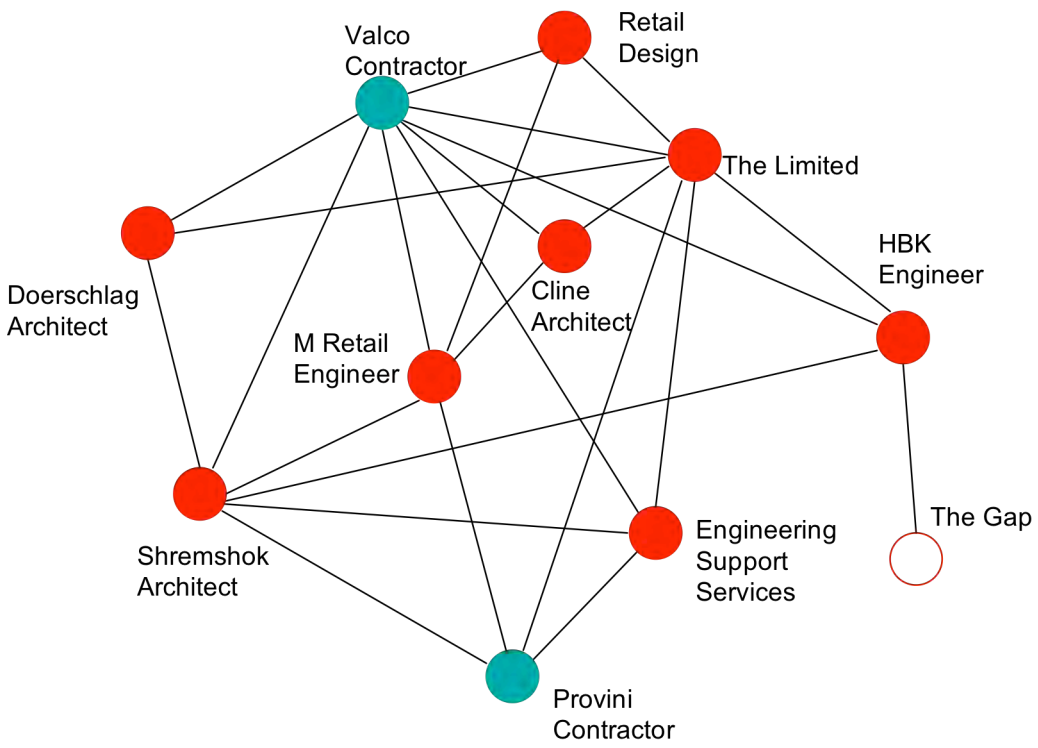


Figure 24. Simplified Limited Company Network

Sociocultural/market model example

To run an effective program, it is vitally important to understand the market and the segments within a market. An example is the retail food sales market in the United States (Figure 25). This model shows how households ultimately obtain their food. While the upper parts of the model are interesting and quite

complex, we focus on the lower parts of the model that are related to retail food sales buildings. A quick check of U.S. DOE Energy Information Administration

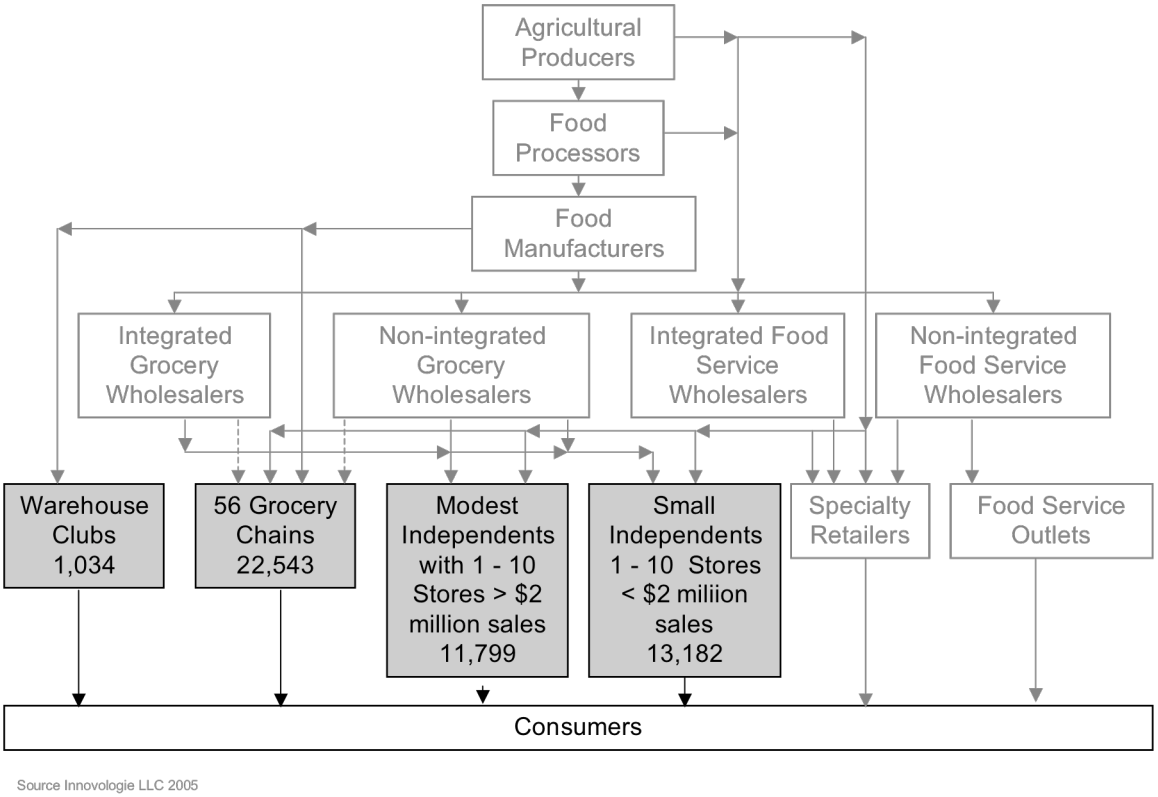


Figure 25. Structure of the U.S. Food Market

(EIA) statistics show that retail food sales (supermarkets) are among the most energy intensive buildings in the commercial building sector. However, there are significant variations in the size of organizations in this market.

In terms of food sales, there are four segments: large warehouse stores, large supermarket chains, modest size independents, and independents. There are three major warehouse chains: SAM's, Costco, and BJ's. There are 56 large supermarket chains that own about 22,500 stores with average store sales of above $2 million. These stores tend to be 30,000 to 60,000 square feet in size but can be as large as 200,000 square feet. These 56 large chains sell 75 percent of the grocer-

ies to U.S. households. The modest sized independents have approximately 11,800 stores, and independents represent approximately 13,200 stores. The modest size independents have ten or fewer stores with most owning one or two stores and store sales that are greater than $2 million but that rarely exceed $4 million. The small independents typically have one or maybe two stores with small footprints of perhaps 10,000 square feet. They are the quintessential "mom and pop" stores.

There are several reasons for needing to understand the structure. First, the large supermarkets are centrally owned and centrally managed. All decisions about what equipment is used and placed in stores are decided centrally. At least figuratively, it takes just 56 contacts to reach these chains. Second, because the margins in the grocery business are very small (3 percent), these chains are looking at any investment that will reduce costs or reduce the increase in costs. They are paying close attention to energy. Third, there are important differences among these chains. Some are merchandizing oriented and some are operationally oriented. Chains that are merchandizing oriented are particularly responsive to customers and worry about aesthetics or barriers (such as doors on cold cases) that may separate the customer from product. Operationally oriented chains worry less about aesthetics and barriers and worry more about execution and cost. Since merchandizing usually trumps operational considerations, products that are merchandizing friendly can take precedence. Finally, these chains are the poster children for the diffusion of innovations model. Their energy managers are constantly looking for new technologies, they investigate the technologies, they do back-of-the-envelope calculations, they install pilot installations, they monitor, and if the technology performs as expected, they will include it in their design prototype and begin installing or retrofitting in their stores. For the most part, these chains have design prototypes that serve as specifications for future construction.

The modest sized independents and the small independents are quite different. There are large numbers of them. They are difficult to contact, because there is no central organization that controls large numbers of stores. Vendors of equip-

ment say the samll independents are not very interested in energy efficient technologies.

As a result of examining the structure and operation of these entities, it appears that the large chains are an important target. They are high energy users. They are few in number. They try to standardize across their stores. They are rational about selecting technologies. Cost, particularly energy cost, is important to them.

9. Summary

The impact evaluation framework described in this document is quite different from evaluation approaches found in other publications about impact analysis. Typically, those materials are technically oriented treatises focusing on measurement and analysis techniques. Those documents and books are quite helpful assuming that one knows and understands the linkages between the outputs and outcomes and what needs to be measured.

This document is less about measurement and analysis techniques and more about providing tools that focus on defining the linkages between outputs and outcomes. The idea is to use sound principles of social science to more clearly identify what needs to be measured, develop better evaluation designs, and better harness existing data collection activities to obtain needed data.

This document starts from the premise that identifying the linkages between outputs and outcomes is one of the most critical and most difficult problems in program design and evaluation. Tools and strategies were presented for increasing understanding of target audience responses to a program so that the responses can be measured more precisely and differentiated from the responses of people and organizations not specifically influenced by the program.

9.1 The Framework

This document presents a broad impact evaluation framework that program managers and/or their contractors can use to develop a well structured impact evaluation. The seven-step process was described in detail in Chapter 3.

1. Identify scope, objective and priorities.
2. Select the types of evaluation to be completed.
3. Select the aspects of deployment induced changes to be evaluated.
4. Identify research questions and metrics.
5. Design the evaluation.
6. Conduct the evaluation.
7. Report and use the results and data.

9.2 Domains

The impact evaluation framework incorporates the fact that deployment programs typically engage one or more of four groups of partners or target audiences: the knowledge community, public entities, manufacturers and other businesses, and end-users. In general, program's intermediate goals are to involve these audiences or partners to:

1. Create, advance, and package market and technical knowledge to make energy efficiency more accessible and implementable (knowledge community).
2. Change the policies, structure, and operations to smooth the advance of energy efficiency (public entities).
3. Create and enhance products, create and align market channels, enhance marketing, and develop installation and support infrastructures (manufacturers and other businesses).
4. Adopt, replicate, institutionalize, and enculturate energy efficient technologies (end-users).

If the players in these four domains respond, then a program can meet its larger goals of reducing energy use and energy intensity, increase use of renewable energy, and reduce emissions while enhancing productivity and global security.

The critical need is to understand partners' and target audiences' responses to program outputs and to understand whether or not these actions of the audiences produce the desired impacts. Using the framework described here helps programs make the linkages transparent, measure the results, and attribute the results to their activities.

9.3 Diffusion of Innovations

The impact evaluation framework is based on ideas drawn from the diffusion of innovations (Rogers, 2003). The diffusion of innovations is a well documented and widely accepted description of how change occurs.

Five basic elements from the diffusion of innovations are:

- The diffusion process — Audiences become aware of technologies or practices; they collect and sort information to persuade themselves whether or not to adopt a technology or practice; they decided whether or

not to adopt; they implement the adoption decision; and they confirm their decision.

- The sociocultural environment — Change takes place in social, political and cultural contexts that can accelerate or impede the advance of technologies or practices. It is important to understand and sometimes change these environments.
- Audience characteristics — Individuals and firms accept innovation at different rates. Individuals and firms are often labeled as innovators, early adopters, the early majority, the late majority, and laggards.
- Product characteristics — Products that are widely adopted typically have relative advantage over other products; are compatible with existing social, cultural, and social systems; have limited complexity; can be tried; and the results of their use easily observed.
- Communications mechanisms — Information can be spread by broadcast (one-to-many methods) or by contagion (one-to-one). One-to-one contact through networks is one of the most powerful ways in which information spreads.

A summary of the diffusion of innovations is found in Chapter 2. Because Rogers' work is so seminal, we encourage users of this document to obtain the book and read it. Program implementers who have done so have found it stimulating and useful.

9.4 Domain-Specific Templates

In Chapter 4, we customized Rogers' work to fit the needs of deployment program managers. We reconfigured Rogers' theory into a more detailed systems style model and represented it as a generic flow diagram which we then linked to a high level program logic model for technology deployment that describes activities, audiences, outcomes, and impacts. We then tailored the generic model to each of the four domains: the knowledge community, public entities, manufacturers and other businesses, and end-users. The result is four domain-specific templates that describe how change can occur in the outcome space of each of the four areas or domains.

We envision that users will identify the domains that are most relevant to the program that they are evaluating. The idea is to then use the templates to describe the expected outcomes for the program in some detail. If program logic already exists, users can compare the templates to the existing logic. Based on the

template, the user may decide that the existing logic is fine, reconfigure the logic, or add to it. If program logic is not available, the user can use the template as a resource to develop one. In recent years, many evaluators and some implementers have begun to realize that a detailed and thorough understanding of anticipated outcomes is needed to make programs more effective and to produce better evaluations.

In constructing the templates, we were somewhat constrained by the fact that we were dealing with programs at a generic level. Thus, the diagrams are intended to be suggestive rather than exhaustive. It is not the intention to provide an exact formulation but to encourage the user to think about the behaviors that a program causes or is assumed to cause and whether those behaviors produce the desired long term outcomes, i.e., energy saving or clean energy behaviors.

To illustrate the point, a program manager and an evaluator can take the diagram for a program that addresses manufacturers or businesses, test it against their program and their experience, and systematically begin to examine how the program works. Some categories will fit, others may have to be modified or added, and some or all are likely to be described in more detailed. These schematics are intended to help program managers and evaluators:

- Identify intermediate effects and their long term impacts in more detail.
- Be more systematic in identifying the elements and linkages that cause program impacts.
- Look at their programs more systemically than they otherwise might.
- Drive this process with the best available social science research.

9.5 Evaluation Questions and Metrics

This document also provides a tool that will help program managers and evaluators identify the important evaluation questions and metrics. The table in Chapter 4 presents a comprehensive set of generic evaluation questions organized around diffusion of innovations concepts. The intent is that users develop their own questions and metrics for appropriate domains by patterning them after the questions in this table. The objective is not to replicate every question and every metric in the table but rather to use it as a kind of checklist from which the user

will identify a small subset of questions that are central to the evaluation being performed.

9.6 Evaluation Design

We have also focused on the issue of evaluation design. This is a much neglected and difficult topic. Design is central to attribution. It is important to design the evaluation so that alternative explanations are eliminated. Identifying a comparison group that is "equivalent" to the participant groups is a key to attribution. This is a tricky problem to which careful attention should be paid. The idea is to help managers and evaluators identify cost effective evaluation designs that will help them take full credit for the effects of their programs.

9.7 Data Collection

There are a variety of reasons why impact evaluations can be costly. A main reason is that evaluations often have to reconstruct data after the fact. This framework emphasizes the potential for using and upgrading existing data collection procedures to support evaluations. Implementers will find that enhanced routine data collection will offer them substantial value and that evaluators will be able to provide better evaluations. We also point to the potential utility of secondary data analysis. We urge program managers to examine the potential for enhanced routine data collection and the potential for using secondary data before turning to primary data collection.

9.8 Some Illustrations

In Chapter 8, we provided some examples of the various concepts related to the evaluation framework. Wherever possible, we used examples from studies that have been conducted for EERE, drawing illustrations from other areas as needed.

10. Acronyms

BBtu	Billions of British Thermal Units – A measure of energy
BTPS	Business Transaction Processing System
CBECS	Commercial Building Energy Consumption Survey
CBO	Community Based Organization
CEUS	California End-User Survey
CFL	Compact Fluorescent Lamp
CFO	Chief Financial Officer
CHP	Combined Heat and Power
CIO	Chief Information Officer
CPS	EERE Corporate Planning System
CPUC	California Public Utilities Commission
DER	Distributed Energy Resources
DOE	Department of Energy
EERE	Office of Energy Efficiency and Renewable Energy
EIA	Energy Information Administration
EPA	Environmental Protection Agency
EPACT	Energy Policy Act
ESA	Energy Savings Assessment
ESAMS	Enterprise Safety Application Management System
ESPC	Energy Service Performance Contracting
FEMP	Federal Energy Management Program
GAO	Government Accounting Office
GPRA	Government Performance and Results Act
HERS	Home Energy Rating System
HDTV	High Definition Television
HVAC	Heating, Ventilation, and Air-Conditioning
IAC	Industrial Assessment Center

ICR	Information Collection Request
ITDP	Industrial Technologies Delivery Program
ITP	Industrial Technologies Program
JOULE	DOE's performance tracking system
kW	Kilowatt
kWh	Kilowatt-hour
LED	Light Emitting Diode
LEED	Leadership in Energy and Environmental Design
LEU	Large Energy User
MEP	Manufacturing Extension Partnership
MP3	An acronym for MPEG layer 3, a compressed audio format
NGOs	Non-Governmental Organizations
NO_x	Nitrogen oxides
NYSERDA	New York State's Energy Research and Development Authority
O&M	Operations and Maintenance
OMB	Office of Management and Budget
PART	Performance Assessment Rating Tool
PIER	California Energy Commission's Public Interest Energy Research Program
PUC	Public Utilities Commission
PWA	Plant Wide Assessment
SO_x	Sulphur oxides
R&D	Research and Development
RBECS	Residential Building Energy Consumption Survey
RCT	Randomized Controlled Trial
RFP	Request for Proposal
TVA	Tennessee Valley Authority
WAP	Weatherization Assistance Program

11. References

ADM Associates and Innovologie, LLC. *Evaluation of the California Residential Appliance Recycling Program 2004 – 2005*. Sacramento, CA: ADM, 2007.

ADM Associates and Innovologie, LLC. *Process Evaluation of 2003 Statewide Codes and Standards Program*. Sacramento, CA: ADM, June 2006.

Aizen, Icek. "The Theory of Planned Behavior." In *Organizational Behavior and Human Decision Processes*. Vol. 50 No. 2, 179-211.

Arnold, Eric. "Evaluating research and innovation policy: a systems world needs systems evaluations," *Research Evaluation*, Vol. 13: 3-17, April 2004.

Backstrom, L., D. Huttenlocher, J. Kleinberg, and X. Lan. "Group Formation in Large Social Networks: Membership, Growth, and Evolution" in the *Proceedings of the 12th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 2006.

Bandura, Albert J. "On Integrating Social Cognitive and Social Diffusion Theories." In Arvind Singhal and James W. Dearing, eds., *Communication of Innovations: A Journey with Ev Rogers*, Thousand Oaks: Sage Publications, 2006. pp 111-135.

Barabasi, Albert-Laszlo. *Linked: The New Science of Networks*. Cambridge: Perseus Publishing, 2002.

Barnes, Harley and Gretchen Jordan. *EERE Guide for Managing General Program Evaluation Studies*. Washington, DC: US Department of Energy, 2006. http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/evaluation_mgmt_guide_final_2006.pdf

Bass, F. M. "A New Product Growth Model for Consumer Durables," *Management Science* 15: 215-227, 1969.

Baumgartner, Robert M. *Customer Response to Load Management: A Survey and Analysis of Utility Studies*, EA-3934, Research Project 2343-3. Palo Alto, California: Electric Power Research Institute, 1985.

Borgatti, Everett, and Freeman. *UCINET 6 for Windows: Software for Network Analysis*. Havard, MA: Analytic Technologies, 1999.

Burby, Raymond J., Nicholas M. Didow, Jr., Mary Ellen Marsden, Charles F. Hanna, Michele D. Bunn, and Stephen A. Johnston. Electric Utility Load Management: Marketing Strategies and Tactics, Totowa, NJ: Rowman and Allanheld, 1985.

Cadmus, Kema-Xenergy. Nexus. *National Analysis of CEE 2002 ENERGY STAR Household Survey*. Boston, MA: CEE, 2003. www.cee1.org/eval/2002-ES_survey_report.pdf.

Cadmus and Kema-Xenergy. *National Awareness of ENERGY STAR Analysis of CEE Household Survey 2003*. Boston, MA: CEE, 2004. www.cee1.org/eval/2003-ES_survey_rep.pdf.

Cadmus and Kema, Nexus. *National Awareness of ENERGY STAR Analysis of CEE Household Survey 2004*. Boston, MA: CEE, 2005. www.cee1.org/eval/2004-ES_survey_rep.pdf.

Dowd, Jeff, Gretchen Jordan, John H. Reed, and Edward Vine. "Are Federal Energy Technology Programs Assessing the 'Magic in the Middle'?" in the *Proceedings of the 2005 International Energy Program Evaluation Conference*. Brooklyn, NY, August 2005.

Eilert, P., N. Horowitz, G. Fernstrom, D. Mahone, and N. Stone. "A Strategic Framework for PGC Planning: Strategic Linkages Between Codes and Standards and Resources Acquisition" in the *Proceedings from the 2002 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy Efficient Economy, 2002.

Eilert, P., J. Livingston, and P. Turnbull. "How to Catch More Flies with Honey and Vinegar: Integrating Voluntary Programs with Codes and Standards in California" in the *Proceedings from the 2004 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy Efficient Economy, 2004.

Fallah, M. Hosein and Elias Aravantinos. "Technology Diffusion and Long Term Forecasting: Application to Growth of Wireless Mobile Services," *AT&T Symposium*.

Hoboken, NJ: Wesley J. Howe School of Technology Management, Stevens Institute of Technology, August 3-4, 2006.

Fishbein, Martin and Icek Aizen. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading: MA: Addison-Wesley, 1975.

Government Accounting Office. *Designing Evaluations*. GAO/PEMD 10.1.4. Washington, DC: GAO, 1991.

Gladwell, Malcolm. *The Tipping Point*. New York: Back Bay Books, 2002.

Granovetter, Mark. *Getting a Job*, Cambridge, MA: Harvard University Press, 1974.

Granovetter, Mark. "The strength of weak ties," *American Journal of Sociology*, Vol. 78, 1973.

Hall, N., et al. *2001 FEMP Customer Survey Study Report*. Oregon, WI: TecMRKT Works, 2002.

Hall, Nick, John Reed, Tom Talerico, Jeff Riggert, Andrew Oh, and Sandia National Laboratories. *2001 FEMP Customer Survery — Process, Market, and Impact Evaluation*. Washington, DC: U.S. Department of Energy, June 2001a.

Hall, Nick, John Reed, Tom Talerico, Jeff Riggert, Andrew Oh, and Sandia National Laboratories. *Executive Summary: 2001 FEMP Customer Survey — Process, Market, and Impact Evaluation*. Washington, DC: US Department of Energy, June 2001b.

Hall, Nicholas P. and John H. Reed. *Evaluation Methods to Quantify Non-Energy Impacts of Vermont's Weatherization Assistance Program*, Burlington, VT: Vermont Weather Assistance Program, January 1998.

Itron, J.J. Hirsh, Kema, and ADM. *California Commercial End-Use Survey - Executive Summary*. Liberty Lake: Washington: Itron, 2006

Johnson, Steven. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York, NY: Touchtone, 2001.

Johnson, Katherine, John Reed, Moria Morrissey, Charles Bailey, and Jeff Riggert. "Building a Better Mousetrap: Better Program Designs for the Retail Food Sub-

market" in the *Proceedings of the 17th National Conference of the Association of Energy Service Professionals*. Las Vegas, NV: AESP, 2007.

Jordan, Gretchen, John Mortensen, John H. Reed, and George Teather. "Developing R&D Logic Models" in the *Proceedings of the 13th International Conference on Technology Management*. Washington, DC: IAMOT, April 2004.

Jordan, Gretchen B. and John H. Reed. "Planning Corporate Performance - The Logical Approach" in the *Proceedings of Linking Performance Measurement to Strategic Planning in Government*. New York: Institute for International Research, New York, NY, 1996.

Kempe, D., J. Kleinberg, and E. Tardos. "Influential Nodes in a Diffusion Model for Social Networks" in the *Proceeding of the 32nd International Colloquium on Automata, Languages and Programming (ICALP)*, 2005.

Kempe, D., J. Kleinberg, and E. Tardos. "Maximizing the Spread of Influence through a Social Network" in the *Proceedings of the 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 2003.

Klovdahl, Alden S. "Social Networks and the Spread of Infectious Diseases: The AIDS Example," *Social Science Medicine*, 21 (11): 1203-1216, 1985.

Lee, Lark, John H. Reed, and Laura Schauer. *Maryland Public Service Commission Electric Universal Service Program Evaluation: Final Evaluation Report*. Baltimore, MD: Maryland Public Service Commission, 2007.

Leskovec, J., A. Singh, and J. Kleinberg. "Patterns of Influence in a Recommendation Network" in the *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2006.

McLaughlin, J. A., and Jordan, G. B. "Logic Models: A Tool for Telling Your Program's Performance Story," *Evaluation and Program Planning*, 22, 65-72. 1999.

Mahajan, Vijay, and Robert A. Peterson. *Models of Innovation Diffusion*. Sage Publications: Newbury Park, CA, 1985.

Martin, M., R. Schmoyer, B. Tonn, and S. Schexnayder. "Study of energy savings generated by clients of the Industrial Assessment Center Program" in the *Proceedings of the 2003 International Energy Program Evaluation Conference*, 2003.

Martin, Michaela and Tykey Truitt. *Estimation of Energy Savings Outcomes from Best Practices Activities, Fiscal Year 2003* (Final Report Draft). Oak Ridge, TN: Oak Ridge National Laboratory, 2004.

McNary, Bill. "ENERGY STAR Sales DATA Collection," A Presentation at the EPA Partners Meeting, D&R International, Ltd., October 2004. www.energystar.gov/ia/partners/downloads/Plenary_D_BillMcNary.pdf

Megdal, Lori M. "Am I Wearing the Wrong Paradigm for My Program Goals?" in the *Proceedings of the International Energy Program Evaluation Conference*. New York, New York: 2005.

Milgram, Stanley. "The Small World Problem," *Psychology Today*, Vol. 1, No. 1,60-67, May 1967.

Molas-Gallart, Jordi and Andrew Davies. "Toward Theory-Led Evaluation: The Experience of European Science, Technology, and Innovation Policies," *American Journal of Evaluation*, Vol. 27, No. 1, 64-82, 2006.

Moore, Geoffrey. *Crossing the Chasm*. New York: Harper Business Books, 1991.

National Cancer Institute. "Behavioral Intentions, Willingness, and Expectations," Health Behavior Constructs, Theory, Measurement, and Research, Cancer Control and Population Sciences. 2007. http://dccps.cancer.gov/brp/constructs/.

Office of Management and Budget. *Guidance for Completing the Program Assessment Rating Tool (PART)*. Washington, D.C., 2005.

PAJEK: Program for Large Network Analysis, http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

PHB Hagler Bailly, Inc. *1998 PG&E Residential Energy Management Services Program: Market Baseline and Market Effects*, Madison, WI, 1999.

Reed, John, Gretchen Jordan and Edward Vine. Overview and Example of *Impact Evaluation Framework for Technology Deployment Programs: An Approach for Quantifying Retrospective Energy Saving, Clean Energy Advances, and Market Effects*. Washington, DC: US Department of Energy, 2007. http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/impact_framework_tech_deploy_2007_overview.pdf.

Reed, John H. and Gretchen B. Jordan, "Using Systems Theory and Logic Models to Define Integrated Outcomes and Performance Measures in Multi-program Settings." *Research Evaluaton*. forthcoming 2007.

Reed, John H., Katherine Johnson, Moria Morrissey, Charles Bailey, and Jeff Riggert. *How They Play and How They Decide: A Study of Commercial Market Decision-Makers*. Rockville, MD, forthcoming 2007.

Reed, John H., Katherine Johnson, Moria Morrissey, Jeff Riggert, and Charles Bailey. "What Happened to My Corner Store? An Examination of the potential for Low Energy or Zero Energy Buildings in the Retail Food Market" in the *Proceedings of the 2006 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington DC: ACEEE, 2006a.

Reed, John H. "A Generic Theory Based Logic Model for Creating a Scientifically Based Program Logic Model: An Expert Lecture," *Annual Meeting of The American Evaluation Association*, Portland, Oregon, November 4, 2006b.

Reed, John H. and Gretchen Jordan. "Generic Models for Federal Program Delivery and Diffusion of Innovation," paper presented at *American Evaluation Association Annual Meetings*. Toronto, October 2005b.

Reed, John H., Katherine Johnson, Jeff Riggert, and Jerry Dion. "The Structure, Ownership, and Energy Use Characteristics of the Retail Sub-market of the US Commercial Buildings Market" in the *Proceedings of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington DC: ACEEE, August 2004a.

Reed, John H., Katherine Johnson, Jeff Riggert, and Andrew Oh. ,*Who Plays and Who Decides*. Washington, DC: U.S. Department of Energy Report: DE-AF26-02NT20528, March 2004b.

Reed, John H. and Andrew D. Oh. "Examining Networks of Building Professionals, Developers, Owners and Contractors in the Commercial Building Sector" in the *Proceedings of the International Energy Program Evaluation Conference,* 2003a.

Reed, John H., Jeff Riggert, and Andrew D. Oh. "Using Business Network Analysis to Target Energy Efficiency Marketing Efforts in Commercial Buildings and to Understand Networks" in the *Proceedings of the Association of Energy Services Professionals*. Boca Raton, FL: AESP, 2003b.

Reed, John H. and Jeff Riggert. *The Use and Energy Savings from Compact Fluorescent Lamps Purchased through the Apartment and Condo Efficiency Services (ACES) Energy Star Products Program*. Madison, WI: Focus on Energy, State of Wisconsin, Department of Administration, Division of Energy, October 6, 2003c.

Reed, John H. and Andrew D. Oh. "Examining Networks of Building Professionals, Developers, Owners, and Contractors in the Commercial Building Sector" in the *Proceedings of the 2003 International Energy Program Evaluation Conference*. Seattle, WA: IEPEC, August 2003d.

Reed, J. H. and Andrew Oh. *The Need for Tools to Support the Federal Energy Management Program: A Report of a Focus Group*. Richland, WA: Pacific Northwest National Laboratories, July 2001.

Reed, J. H., Tom Talerico, Andrew Oh, and Jeff Riggert. *The Market Effects of the Energy Training Center Stockton*. San Francisco: Pacific Gas and Electric Company, 2000.

Reed, John H., Andrew D. Oh, and N. P. Hall. *Market Progress Report, Silicon Growing Facilities No. 1*. Portland, OR: Northwest Energy Efficiency Alliance, 1999a.

Reed, John H, Nicholas P Hall, and Andrew Oh. *An Impact Evaluation of the Missouri Gas Energy Low-Income Weatherization Pilot Program*. Kansas City, MO: Missouri Gas Energy, May 1999.

Reed, John H., Nicholas P. Hall, Andrew Oh, Paul Hoover, and John Oh. *An Impact Evaluation of Vermont's Weatherization Assistance Program*. Burlington, VT: Vermont Weather Assistance Program, December 1997.

Renger, Ralph. "Consequences to Federal Programs When the Logic-Modeling Process Is Not Followed With Fidelity," *American Journal of Evaluation*, Vol. 27: 452-463, December 2006.

Rheingold, Howard. *Smart Mobs: The Next Social Revolution*. Cambridge, MA: Persus Books, 2002.

Rogers, Everett. *Diffusion of Innovations Fourth Edition*. New York: New York Free Press, 1995.

Rogers, Everett. *Diffusion of Innovations Fifth Edition*. New York: New York Free Press, 2003.

Romm, Joseph, "The Hype About Hydrogen", *Issues in Science and Technology*, Dallas: University of Texas, April 2004.

Rosen, Emanuel. *The Anatomy of Buzz: How to Create Word of Mouth Marketing*. New York: Doubleday-Currency, 2000.

Ruegg, Rosalie, and Gretchen Jordan. *Overview of Evaluation Methods for R&D Programs*. Washington, DC: US Department of Energy, 2007. http://www.eere.energy.gov/ba/pba/km_portal/docs/pdf/2007/evaluation_methods_r_and_d.pdf.

Schweitzer, M. and J. Saulsbury. *An Examination of Record-Keeping Practices by Selected DOE Energy Efficiency and Renewable Energy Deployment Programs*, Oak Ridge National Laboratory, Oak Ridge, TN, 2005.

Scott, John. *Social Network Analysis: A Handbook*. 2nd ed. London: Sage Publications, 1991, 2000.

Taylor, Mark C. *The Moment of Complexity: Emerging Network Culture*. Chicago: Chicago University Press, 2001.

Taylor, Jennifer. "LRC explores basic performance parameters of LEDs," Troy, New York: Lighting Research Center, Rennselaer Polytechnic Insititue, 2007. http://www.lrc.rpi.edu/resources/newsroom/enews/Oct06/Research21.html

TecMRKT Works and Sandia National Laboratories. *2001 FEMP Customer Survey Study Report*. Madison, WI, 2002.

TecMRKT Works. *The California Evaluation Framework*. Madison, TecMRKT Works, 2004.

Tornatzky, L. G. and J. J. Klein, "Innovation Characteristics and Innovation Adoption – Implementation: A Meta-analysis of Findings." *IEEE Transactions of Engineering and Management*, EM-29:28-45, 1982.

U.S. Department of Energy. *PEER Review Guide*. Washington, D.C.: US Department of Energy, 2004. www1.eere.energy.gov/ba/pdfs/2004peerreviewguide.pdf

Vine, Ed, Gretchen Jordan, John H. Reed, and Jeff Dowd. "An Inside Look at a U.S. Department of Energy Impact Evaluation Framework for Deployment Programs" in the *Proceedings of the 2006 ACEEE Summer Study on Energy Efficiency in Buildings*. Washington, DC: ACEEE, 2006.

Watts, Duncan J., and Steven Stogratz. "Collective dynamics of 'small-world' networks," *Nature*, 393, 440-442, 1998.

Watts, Duncan J. *Six Degrees: The Science of a Connected Age*. New York: Norton, 2003.

Whyte, William H., Jr. "The Web of Word-of-Mouth," *Fortune*, November 1954.

12. Index

A

activities, iv, v, viii, 1-4, 1-5, 1-6, 1-8, 2-6, 3-3, 3-4, 3-7, 3-9, 3-10, 3-13, 3-14, 3-15, 3-26, 4-1, 4-4, 4-5, 4-6, 4-22, 4-24, 4-26, 5-2, 6-15, 6-16, 7-1, 7-4, 7-6, 7-10, 8-1, 8-9, 9-1, 9-2, 9-3
  high level, 4-2
administrative costs, 3-15
adoption, vi, 1-4, 2-1, 2-3, 2-4, 2-5, 2-6, 2-9, 2-10, 2-11, 3-16, 3-18, 3-23, 4-3, 4-10, 8-8, 9-3
  adopter characteristics, 3-18, 5-1, 8-5
  adopter types, 2-8, 4-6, 5-3, 5-4
  early adopters, vi, 2-8, 2-9, 3-17, 3-18, 4-6, 4-10, 4-19, 5-2, 5-3, 5-11, 9-3
  early majority, vi, 2-8, 2-9, 3-17, 3-18, 4-10, 5-3, 5-11, 9-3
  funding, 4-2, 4-3, 4-5
  laggards, vi, 2-8, 2-9, 5-3, 9-3
  late majority, vi, 2-8, 2-9, 4-19, 5-3, 9-3
  promoting, 4-2, 4-3, 4-5
  S-curve, 2-6, 8-5
advertising, 2-5, 6-16
  bill inserts, 2-5, 3-18
  brochures, 3-10, 4-3
  e-mail blasts, 3-19
  flyers, 3-18
  Internet, 2-7, 3-19
  list serves, 3-19
  magazine articles, 2-5
  mass mailings, 2-5
  mass media, 2-5
  newspapers, 2-2, 3-18
  radio, 3-18
  television, 2-2, 2-7, 3-18, 8-6
  trade publications, 3-18, 4-15
  websites, 2-5, 3-10, 3-19, 4-3, 4-4, 4-22, 4-26
aided recognition, 8-12
alternative fueled vehicles, 4-3
appliance
  air conditioners, 8-7
  Clothes washers, 8-7
  efficiency, 2-4
  heat pumps, 3-20, 4-10
  HVAC, 6-3, 10-1
  market, 2-4
  recycling programs, 6-7
  refrigerators, 2-6, 6-7, 8-7
audience
  behaviors, iv
audits, 1-6, 3-10, 4-22, 6-8, 7-1, 8-2
  audit program, 4-13, 4-22
  energy, 3-7, 4-9
Awareness, 2-2, 2-3, 4-15, 4-19, 4-21, 4-23, 5-4, 6-15, 7-3, 11-2
  brand, 8-12

B

Bandura, 3-18, 11-1
barrier, 2-4, 3-23, 3-24, 5-4, 5-4, 8-3, 8-15
baseline, 1-6, 3-13, 3-15, 5-8, 6-2, 7-4, 7-5, 7-6
behavior
  planned, 2-12
  pre-program, 6-2
  sustained, 3-9, 3-23, 8-2
behavioral
  change, 1-3
  data, 7-4
  intentions, 2-12
  persistence, 6-6
  questionnaire, 7-5
bell curve, 8-5
bias, 4-9, 6-14
broadcast, vi, 2-2, 2-5, 2-6, 2-9, 3-18, 3-19, 4-6, 5-10, 9-3
budgeting, 4-18
building
  baseline, 2-11
  codes, 2-10, 4-3, 4-21
  commercial, 1-3, 3-7, 4-10
  commissioning, 4-14
  simulation modeling, 7-9
business
  lines, 4-17
buy downs, 4-21

C

California Evaluation Framework, 6-1, 11-9
call center, 3-10, 4-4
capital, 3-16, 3-18, 3-23, 4-8, 4-9, 4-14, 4-18, 4-24, 5-6, 7-5, 8-1
  investment, 3-16, 8-1
case study, iv, 2-12, 4-3, 4-4, 7-3
CBECS, 7-8, 10-1
CFLs, 2-3, 2-5, 3-12, 3-13, 3-20, 4-10, 6-6
CFO, 4-9, 10-1
change
  non-program driven, 6-4

change agent, 2-9, 4-10
channels, 2-8, 2-9, 3-10, 3-12, 3-13, 3-18, 3-19, 4-18, 4-19, 5-10, 5-12
codes and standards, 4-11, 4-15, 4-17, 4-18, 4-21, 4-26, 5-9
colleges, 3-11, 3-16, 4-5
comfort, 2-9, 4-14, 6-10, 7-4
communication, 2-1, 2-5, 2-8, 3-23, 4-19
  broadcast, vi, 2-2, 2-5, 2-6, 2-9, 3-18, 3-19, 4-6, 5-10, 9-3
  contagion, vi, 2-6, 2-7, 2-9, 3-18, 3-19, 4-6, 5-10, 9-3
  model, 2-5
  network, vi, 2-5, 2-6, 2-7, 3-24, 4-18, 4-22, 5-9, 5-10, 8-12, 9-3, 11-5, 11-9
Community Action Agency, 6-12
competitive advantage, 4-17
competitive environment, 3-16
complexity, vi, 1-1, 2-11, 3-4, 3-18, 3-20, 3-21, 4-8, 4-13, 4-17, 4-24, 5-13, 8-3, 8-4, 8-5, 9-3
confidence level, 6-1
Congress, 1-6
consultant, 3-4, 3-8
consumer, 3-21, 4-24, 5-12, 6-7
contagion, vi, 2-6, 2-7, 2-9, 3-18, 3-19, 4-6, 5-10, 9-3
content development, 3-9
contracts, 4-21, 6-8
cool roofs, 4-10
corporate culture, 3-16
cost effectiveness, iii, 3-14
CPS, 10-1
CPUC, 10-1
credibility, 3-26, 4-23
cross fertilization, 8-13
cultural, vi, 2-4, 2-5, 2-7, 2-10, 4-10, 4-21, 7-9, 9-3
  acceptability, 4-21
cycle times, 4-14

D

data
  cost, 3-14, 3-15
  gaps, 6-1
  pre-existing, 7-1
  secondary sources, 7-1
  utility billing, 7-9
data collection, iv, viii, 1-8, 1-10, 5-1, 7-1, 7-6, 7-7, 7-9, 7-10, 9-1, 9-5
  post-hoc, 7-1
  primary, 7-1, 9-5
  procedures, viii, 9-5
  strategies, 1-10
decision makers, 1-3, 1-4, 2-4, 4-10, 5-9, 5-8
defensibility, 6-1
delivery
  channels, 3-3, 3-10, 3-11, 3-12, 5-10, 5-12
demand, 3-19, 4-19, 6-16
demographics, 8-7
Department of Human Services, 6-12
deployment, v, vii, 1-1, 1-4, 1-5, 1-6, 1-8, 2-4, 3-1, 3-7, 3-11, 3-14, 3-20, 3-25, 4-1, 4-2, 4-3, 4-4, 4-5, 4-6, 4-10, 4-11, 7-1, 8-11, 9-1, 9-2, 9-3
  induced changes, iv
design
  cross-sectional, 7-6
  green, 4-21, 4-22
  panel, 7-6
  quasi-experimental, 6-11
developers
  large, 2-5
diffusion curve, 2-6
Diffusion of innovation, vi, 2-1
diffusion process, vi, 2-1, 3-23, 8-8, 9-2
distribution systems, 4-17
DOE, i, 1-8, 3-25, 4-4, 6-8, 7-3, 7-10, 8-14, 10-1, 10-2, 11-8
domain, v, vii, 1-4, 1-5, 1-6, 1-7, 2-4, 3-3, 3-7, 3-8, 3-16, 3-17, 4-1, 4-11, 4-15, 4-19, 4-21, 4-22, 4-24, 4-26, 5-1, 6-2, 6-3, 6-7, 9-2, 9-3, 9-4
  business, 1-6, 1-7, 1-8, 4-15, 4-18, 6-3, 6-7
  public entities, 4-19
domain-specific, vii, 9-3
double counting, 6-7, 6-8

E

EERE, i, v, viii, 1-1, 1-6, 1-7, 1-10, 3-9, 3-10, 3-13, 3-19, 3-23, 3-25, 3-26, 4-1, 4-2, 4-4, 4-5, 4-6, 4-8, 4-11, 4-22, 6-1, 7-1, 7-4, 7-9, 7-10, 8-2, 8-9, 9-5, 10-1, 11-1
  Steam System Tool, 4-8
EERE Peer Review Guide, 1-10
effects
  direct, 3-23, 4-10, 6-5
  externally induced, 4-9
  indirect, iii, 1-1, 6-5
  intermediate, 9-4
  non-program, 3-12
  primary, 6-5
  program, iii, 6-4, 6-5, 6-9
  secondary, 6-5
  sustained, 6-5, 6-6
efficient
  buildings, 4-21
  practices, 1-3

technologies, 1-3, 4-18, 5-7
EIA, 7-10, 8-14, 10-1
Eilert, P., 4-18, 11-2
electronic ballasts, 8-5
emissions, v, 3-6, 4-22, 7-3, 7-4
reduction, 1-3, 9-2
emulation, 1-6, 3-9, 3-23, 3-24, 4-10, 4-22, 5-8, 6-5, 6-7, 8-2
energy
consumption, 4-9, 6-2, 6-13, 7-4, 8-5
intensity, v, 7-5, 7-6, 9-2
renewable, 2-1, 4-1, 4-5, 9-2
savings, 1-1, 1-9, 3-7, 3-8, 3-14, 4-22, 7-5, 7-7, 8-1, 8-2, 8-4, 8-10, 11-5
use, v, 1-3, 2-11, 5-2, 5-8, 6-2, 6-9, 6-14, 6-15, 7-8, 7-9, 8-4, 8-9, 8-16, 9-2
energy efficiency, v, 1-5, 1-9, 2-1, 2-3, 2-12, 3-6, 3-18, 3-20, 4-1, 4-3, 4-4, 4-5, 4-11, 4-13, 4-15, 4-19, 4-21, 4-22, 5-5, 5-7, 5-8, 6-4, 6-5, 6-6, 6-9, 8-1, 8-3, 8-5, 8-9, 8-13, 8-16, 9-2
energy saving
retrospective, 1-1
Energy Service Performance Contracting, 6-11, 10-1
Energy Star, 3-8, 3-22, 4-3, 4-24, 6-15, 6-16, 8-7, 8-11, 8-12, 11-7
engineering firms, 3-17, 8-13
entities
governmental, 4-19
public, iii, v, vii, ix, 1-4, 1-9, 3-15, 4-3, 4-5, 4-8, 4-11, 4-19, 4-21, 4-22, 5-1, 6-5, 9-2, 9-3
ESPC, 6-11, 6-12, 10-1
evaluation
activities, 1-10, 3-9, 3-10, 3-16, 4-2, 4-7, 5-1, 9-3
benefit-cost, 3-14
community, 2-12
contractor, iii, 1-1, 3-26
cross-program, 3-10
design, iv, viii, ix, 1-2, 1-8, 1-10, 3-25, 6-1, 6-2, 7-6, 9-1, 9-5
process, 3-15
questions, vii, 1-3, 1-10, 6-2, 9-4
scope, 6-1
evaluators, iii, iv, vii, viii, 1-1, 1-2, 1-6, 1-7, 1-8, 3-4, 3-6, 3-19, 3-20, 3-23, 3-25, 3-26, 3-27, 4-1, 5-1, 6-15, 6-17, 9-4, 9-5
independent, 1-1
external factors, 3-6, 4-2, 6-3, 8-2

F

facility engineer, 3-4, 3-24, 4-9
feedback, 1-3, 1-6, 2-1, 3-4, 4-4
FEMP, 4-4, 4-13, 4-23, 6-11, 6-12, 8-8, 8-9, 8-10, 10-1, 11-3, 11-9
financial assistance, 1-6
Flex Your Power, 6-7
free riders, 1-9, 6-4
FW Dodge Players data, 8-12

G

GAO, 6-1, 11-3
GAO's Designing Evaluations, 6-1
gatekeeper, 2-8
General Program Evaluation Guide, 3-13
global security, v, 9-2
goals, iii, iv, v, 1-1, 1-2, 1-3, 3-3, 3-9, 4-19, 7-2, 8-11, 9-2
energy efficiency, 4-11
long-term, 1-3, 4-5
program, 1-2, 3-3, 3-6
GPRA, 1-8, 3-6, 10-1
grants, 4-4, 4-21, 7-2
green design, 4-21, 4-22
grocery chains, 2-9

H

Hall, Nick, 3-14, 6-11, 6-13, 8-8, 8-10, 8-11, 11-3, 11-7, 11-8
homebuilders, 2-4
households, 1-3, 4-10, 5-8, 6-10, 6-12, 6-13, 8-13, 8-15
human resource, 4-17
human services, 2-1

I

IAC, 3-4, 8-9, 8-10, 10-1
ICR, 7-9, 10-2
impact
net, 3-12, 3-13, 6-2
impact evaluation, iii, iv, v, vi, vii, 1-1, 1-2, 1-8, 1-9, 1-10, 2-7, 3-1, 3-3, 3-6, 3-7, 3-12, 3-14, 3-15, 3-23, 3-25, 4-6, 4-13, 6-2, 6-8, 6-17, 7-8, 9-1, 9-2
impacts
desired, v, 1-3, 1-5, 1-10, 9-2
gross, 3-12, 6-2
long-term, 9-4
program, vii, 9-4
implementation, iv, 1-3, 2-2, 2-3, 2-4, 2-12, 3-13, 3-15, 3-19, 3-22, 3-26, 4-8, 4-9, 4-13, 4-14, 4-21, 4-24, 5-4, 5-6, 5-6, 5-7, 6-11, 7-2, 7-3, 7-4, 8-9, 8-10
implementers, 1-6, 2-6, 3-26, 3-27, 5-1, 9-3, 9-4
industrial plants, 1-3, 7-1, 8-1

inert gas, 4-14
Information Collection Request, 7-9, 10-2
information seeking, 2-2, 4-15, 7-3
infrastructure, 1-4, 3-9, 4-3, 4-5, 5-9, 6-2, 8-7
  building program, 4-2
  governmental, 4-3, 5-9
  support, v, 4-5, 9-2
innovators, vi, 2-8, 2-9, 3-17, 3-18, 4-6, 4-10, 4-19, 5-2, 5-3, 5-11, 9-3
installation, v, 1-9, 2-5, 3-10, 3-12, 3-23, 4-5, 5-9, 6-3, 8-4, 8-15, 9-2
installers, 3-11, 3-21, 4-18, 5-9, 5-10, 5-13, 6-4
institutionalize, v, 4-5, 9-2
inventory systems, 4-18
investing, 3-18, 4-24
ITDP, 3-4, 3-6, 3-8, 3-9, 4-4, 4-10, 4-13, 4-15, 4-23, 6-5, 7-1, 7-4, 7-5, 8-1, 10-2
  Allied Partner program, 4-4
  Industrial Assessment Center, 3-4, 7-1, 8-9, 10-1, 11-5
Itron, J.J., 7-9, 11-3

J

Jordan, Gretchen, 1-1, i, 4-1, 4-2, 4-6, 4-7, 4-11, 4-12, 4-15, 4-16, 4-20, 4-25, 6-3, 11-2, 11-4, 11-6, 11-9
JOULE, 1-8, 7-2, 10-2
jurisdiction, 4-21, 4-22

K

KEMA, 8-11, 8-12
knowledge
  assembling, 4-3
  communities, 4-24
  community, v, vii, 9-2, 9-3
  creation, 1-4, 3-7
  dissemination, 1-4, 4-3, 4-23
  domain, ix, 1-6, 1-9, 3-4, 4-22
  packaging, 6-3
  sector, 4-11
  technical, v, 4-3, 4-5, 9-2
  workers, 1-3, 1-4

L

laboratories, 4-5, 4-22
LED, 3-23, 4-17, 8-4, 10-2
Lee, Lark, 6-12, 11-4
lighting, 2-11, 3-17, 3-23, 3-24, 4-10, 4-13, 4-17, 6-6, 8-4
load management, 4-9
loans, 4-21
local codes, 1-4, 2-5, 3-16
local norms, 2-5
local regulations, 3-16
logic model, 1-8, 3-1, 3-3, 3-4, 3-6, 3-9, 4-1, 4-2, 4-5, 4-6, 4-14, 5-1, 6-2
low income, 1-3, 4-4, 6-12, 6-13

M

maintenance practices, 4-8, 4-9
manager
  energy, 4-11, 4-22, 5-7, 8-15
  facility, 8-1
  program, iii, viii, 1-7, 3-3, 3-6, 3-20, 3-21, 3-25, 4-13
manuals, 4-3
market
  channels, v, 4-5, 9-2
  effects, 1-1, 1-3, 1-9, 3-13, 3-14, 3-15, 3-17, 4-6, 4-18, 5-1, 7-4, 7-5, 7-9, 8-3
  gaps, 4-21
  leadership, 4-17, 4-19
  model, 2-4, 8-13
  pilots, 3-18
  research, 4-15
  share, 4-19, 7-9
  transformation, 1-9, 2-7, 2-9, 8-10
marketing, v, 2-1, 2-6, 3-15, 4-5, 5-2, 5-3, 5-4, 5-12, 6-7, 7-1, 9-2
markets
  commercial submarkets, 3-17
Martin, M., i, 8-1, 8-9, 8-10, 11-3, 11-5
McLaughlin, J.A., 4-1, 11-4
measurement, iv, 1-7, 2-3, 3-25, 6-2, 6-5, 7-6, 9-1
Megdal, Lori M., i, 2-12, 11-5
metering
  net, 1-4, 2-5, 4-3
  on-site, 7-9
method
  quasi-experimental, 6-11
metrics, iv, vii, 1-2, 1-3, 1-10, 3-1, 3-25, 5-1, 9-1, 9-4
module, 1-2, 3-1
moisture control, 4-24
multifamily, 4-9

N

network analysis, 2-7, 3-24, 8-12
NGOs, v, 1-4, 3-7, 3-11, 4-5, 4-11, 10-2
non-participant, 5-2, 5-4, 5-4, 5-5, 5-6, 5-6, 5-7, 6-5, 6-11, 6-12, 6-13, 6-15, 7-6, 7-7

O

Oak Ridge National Laboratory, 11-5, 11-8
observability, 2-11, 2-12, 3-20, 4-24, 8-4
OMB, 1-6, 1-8, 3-6, 6-8, 6-9, 6-10, 7-9, 7-10, 7-10, 10-2, 11-5
operating
  costs, 8-9
  practices, 4-8
organizational connections, 3-4
organizations
  governmental, 3-15
outcomes
  long-term, iii, vii, 9-4
output, iii, iv, 1-2, 1-5, 1-6, 1-9, 1-10, 3-3, 3-9, 3-11, 3-12, 3-13, 4-1, 4-2, 4-3, 4-5, 4-6, 7-1, 8-1, 8-2, 9-1
outputs
  program, iii, v, 1-6, 3-4, 4-6, 7-2, 9-2
outreach, 4-3, 4-4
outsourcing, 3-6

P

parameters, 2-6
PART, 1-8, 3-6, 6-8, 10-2, 11-5
participation, 3-17, 3-19, 3-25, 5-4, 5-6, 6-11, 7-6, 8-8
paybacks, 8-4
penetration levels, 2-6
perceived behavioral control, 2-12
performance, 1-8, 3-6, 4-4, 4-11, 4-21, 4-22, 5-8, 5-12, 5-14, 6-8, 7-10, 8-4, 10-2
performance contracting, 4-22
persuasion, 2-1, 2-4, 2-12, 3-22, 4-6, 4-7, 4-13, 4-23, 5-4, 5-5, 6-11, 7-2, 7-3
Plant Wide Assessments, 4-4
policy maker, 1-3
policy planners, 1-1
political, vi, 2-1, 2-5, 4-21, 9-3
  cost, 4-21
  science, 2-1
pollution, 1-3, 7-3, 7-4
population
  equivalent, 6-12, 6-13
product characteristics, 2-1, 2-12, 3-20, 4-17, 4-21, 5-1, 5-3, 5-12, 8-3, 8-4, 8-5
  compatibility, 2-10, 3-20, 3-21, 4-8, 4-13, 4-21, 4-23, 5-12, 8-4
  complexity, vi, 1-1, 2-11, 3-4, 3-18, 3-20, 3-21, 4-8, 4-13, 4-17, 4-24, 5-13, 8-3, 8-4, 8-5, 9-3
  credibility, 3-26, 4-23
  defective, 4-14
  defensibility, 6-1
  model, 8-3
  observability, 2-11, 2-12, 3-20, 4-24, 8-4
  trialability, 2-11, 2-12, 3-20, 8-4
product line, 3-16, 4-17, 4-19
production facility, 4-18
production personnel, 4-18
productivity, v, 9-2
professional organizations, 3-16
professional service, v
profit, 4-15, 4-19
program
  accomplishments, iv
  activities, v, 1-2, 1-5, 2-4, 3-3, 3-13, 3-25, 4-3, 4-4, 5-6, 6-2, 7-2, 7-3
  effectiveness, iii, 6-9
  elements, iii, 1-9, 3-13
  logic, 1-7, 3-3, 9-3
  planners, iii, 1-1, 1-6
programs
  information, 4-22
  intervention, 2-1
  maintenance, 4-22
  pilot, 3-18, 3-22, 4-22, 8-4
  resource acquisition, 1-9, 3-15, 6-6
  social networking, 3-19
  weatherization, 3-14, 6-13
prototype, 8-15
public
  facilities, 1-3
  health, 2-1, 3-11
  institutions, 1-4, 1-5
  partnerships, 4-3
  policies, 1-4, 1-5
publications, iv, 2-2, 3-10, 4-1, 4-4, 4-22, 4-26, 7-2, 8-2, 9-1
publicity
  high areas, 6-15
  low, 6-16
PUC, i, 6-12, 10-1, 10-2
purchasing practices, 4-4
PWAs, 8-1

Q

quality control, 3-26
questionnaire
  behavioral, 7-5

R

R&D, 1-10, 4-1, 4-23, 5-3, 5-8, 5-12, 10-2, 11-4
randomized controlled trials, 6-9
RBECS, 7-8, 10-2
RCTs, 6-8, 6-9, 6-10, 6-11, 6-14
reduced energy use, 4-14

Reed, John H., 1-1, i, 2-5, 2-7, 2-11, 3-12, 3-14, 3-17, 3-23, 4-1, 4-2, 4-6, 4-7, 4-11, 4-12, 4-14, 4-15, 4-16, 4-19, 4-20, 4-22, 4-23, 4-25, 6-3, 6-13, 7-8, 8-3, 8-12, 11-2, 11-3, 11-4, 11-6, 11-7, 11-8, 11-9
refrigeration, 3-17, 8-5
relative advantage, vi, 2-10, 2-12, 3-20, 3-21, 4-8, 4-13, 4-15, 4-21, 4-23, 5-11, 5-12, 8-4, 9-3
reliable sources, 4-23
renewable, 1-3, 1-5, 2-1, 4-1, 4-3, 4-5, 9-2
renewable sources, 1-3
replication, 1-6, 3-9, 3-23, 3-24, 4-10, 6-4, 6-5, 6-7, 8-1, 8-2, 8-10
  external, 8-9
  internal, 8-9
reporting, 1-3, 3-25, 4-2, 4-4, 8-1, 8-2
research
  accuracy, 6-1
  in-house, 4-15
  lab, 4-18
  questions, iv, 9-1
resource
  allocation, 3-9, 3-10
resources, 1-6, 2-9, 3-1, 3-3, 3-9, 3-10, 3-27, 4-4, 4-23, 6-1, 6-5
retail food sales, 7-8, 8-3, 8-13
retrofit, 4-14, 4-22, 8-15
reviewing, 4-2
RFP, 3-26, 10-2
risk averse, 2-9
Rogers, Everett, i, vi, ix, 1-7, 2-1, 2-2, 2-8, 2-9, 3-18, 4-1, 4-6, 4-10, 4-11, 5-1, 8-10, 9-2, 9-3, 11-1, 11-8
R-value, 6-6

S

sampling, 2-11, 3-27, 6-1, 6-12, 6-13, 6-14, 7-6, 7-8, 8-12
  comparison group, 6-11, 6-12, 6-13, 6-14, 6-17, 7-1, 7-4, 7-6, 7-7, 9-5
  control group, 6-8, 6-9, 6-10, 6-11
  random assignment, 6-10
  self-selected group, 6-12
  treatment group, 6-11, 6-12, 6-14, 7-6
  zip codes, 6-13
Saulsbury, J, 7-1, 7-3, 7-4, 11-8
savings
  gross, 3-12, 7-1
  net, 7-1, 8-2
Schweitzer, M., i, 7-1, 7-3, 7-4, 11-8
scientifically, iii, 1-1, 1-7, 1-8
scope, iv, 1-2, 1-7, 3-1, 3-6, 6-2, 9-1
S-curve, 2-6, 8-5
sector
  private, 4-21, 7-2
services
  manufacturing extension, 3-16
silicon production furnaces, 4-14
social change, vi, 1-7, 1-9, 2-1
social sciences, iv, vii, 1-8, 9-1, 9-4
social systems, vi, 9-3
socio-cultural environment, vi, 3-16, 3-17, 5-9, 9-3
socio-cultural factors, 3-16
sociology, 2-1
software, 2-10, 3-4, 4-1, 4-3, 4-4, 4-22, 4-23, 4-26, 6-8, 7-2, 7-5, 8-2
solar technology, 3-16
spillover, 1-9, 3-9, 3-23, 4-22, 6-5
spin-off, 5-8, 6-5, 6-7, 8-2, 8-9, 8-10
  measures, 8-10
stage theory, 2-3
stakeholder, iii, 1-3, 3-6, 3-9, 3-10, 5-9
standards
  emerging, 4-3
standards and codes, 4-11, 4-15, 4-17, 4-18, 4-21, 4-26, 5-9
steam systems, 3-4, 7-5
stocking practices, 3-16
study, benefit-cost, 3-14
sub-activities, 3-9, 4-3
supermarket industry, 3-17
systems integration, 3-17

T

target actors, 1-6, 3-7
target audience, iii, iv, v, 1-1, 1-2, 1-6, 1-7, 1-9, 3-4, 3-19, 3-22, 4-1, 4-4, 4-5, 4-6, 4-7, 4-8, 4-10, 5-2, 5-10, 6-2, 6-13, 6-17, 9-1, 9-2
target population, 2-4, 6-15
tax incentives, 4-21
technical assistance, 3-9, 3-10, 4-4, 8-11
technical support, 4-3, 4-4
technological persistence, 6-6
technology
  delivery activities, 1-1
  demonstrations, 4-22, 6-15
  dissemination, 4-3
  distributors, 3-11, 4-5, 4-18, 5-9
  energy efficient, v, 4-1, 4-3, 8-16, 9-2
  evolution, 2-6
  new, 1-3, 2-6, 3-18, 4-8
  renewable, 1-3, 1-5, 4-1, 4-3, 4-5
  retailers, 1-3, 2-11, 3-7, 3-8, 3-11, 3-17, 4-5, 4-10, 4-15, 5-9, 5-10, 6-4
  turn key, 3-17
  viable market, 1-3
template, iii, iv, vii, 1-9, 4-14, 6-2, 9-3, 9-4

testimonial, iii
theory of change, 4-1, 4-5, 4-6, 4-11, 4-15
trade association, 4-15
training, 1-6, 3-4, 3-7, 3-8, 3-9, 3-10, 3-25, 4-3, 4-4, 4-13, 4-14, 4-18, 4-22, 5-13, 6-4, 6-8, 6-11, 7-3, 7-4, 7-5, 7-6, 7-7, 8-8
    delivery, 3-9
    joint programs, 3-10
    programs, 3-4, 3-7, 3-9, 4-13, 6-4, 6-8, 6-11, 7-4
trialability, 2-11, 2-12, 3-20, 8-4
TVA, 3-20, 10-2

U

unaided recognition, 6-16
unions, 2-5
universities, 3-11, 3-16, 4-5, 4-22
utility bills, 2-11, 6-12

V

Vine, Ed, 1-1, i, 4-6, 4-15, 11-2, 11-9

W

WAP, 4-4, 6-10, 10-2, 11-3, 11-8
weighting, 6-13, 6-14, 7-6
wind, 4-3