USING A MIX OF PROGRAM THEORY EVALUATION, AND COMPARISONS AS EVIDENCE FOR CAUSALITY

Lori Megdal, Ph.D., Megdal & Associates
Mark Coleman, Jennifer Ellefsen, Larry Pakenas, and Helen Kim,
New York State Energy Research and Development Authority (NYSERDA)
Scott Albert, GDS Associates, Inc.

This paper will: (1) provide an overview of the causality concept, as defined by the evaluation profession, as well as by the New York State Energy Research and Development Authority (NYSERDA), (2) discuss the causality assessment methodologies being used by NYSERDA, and (3) provide early evidence of causality, as an example of the NYSERDA approach, for **New York Energy $mart**[SM] programs that are impacting the residential appliance and lighting markets.

Defining the Causality Concept

Prior evaluation efforts of the **New York Energy $mart**[SM] program have examined the program interventions as they were designed to achieve specific policy goals and target market barriers, process effectiveness of the program implementation, and current measurements of direct and long-term outcomes from this implementation. Lying beneath these views is the notion that if the program interventions are performed well and designed to meet unfilled customer needs or market barriers, then the program will clearly contribute to outputs that directly cause outcomes, measured in both near and long-term horizons.[a] There are two elements within the latter part of this statement: (1) that the direct and long-term outcomes occur; and (2) that it is the **New York Energy $mart**[SM] program that caused these outcomes, rather than a general trend or other factors. This second element is the focus of the causality assessment and the topic of this paper.

Basically defined, *"Causal attribution is the claim that x caused y. More specifically, in evaluation the claim that the program was responsible for the observed effect."*[b] An examination of causality is an important component of the **New York Energy $mart**[SM] evaluation effort because it can help to validate the existence of program interventions, justify

[a] Outputs consist of program activities and include number of participating customers and trade allies, incentives distributed, and informational activities. Near-term outcomes consist of benefits derived from program participants and include energy savings, emission reductions, and leveraging of program funding. Long-term outcomes include market transformation indicators such as changes in attitudes and behaviors with respect to energy efficiency, improvements in infrastructure to support energy efficiency, change in market share of energy efficient products, and changes in manufacturing standards and regulatory codes.

[b] Carol H. Weiss. 1998. *Evaluation: Methods for Studying Program and Policies*, Prentice Hall, Upper Saddle River: New Jersey, Glossary, pg. 328.

the expenditure of public benefits funds, and assist policy-makers in decisions regarding energy efficiency, low-income and research and development programs. The potential results of causality can have impacts beyond proving "x caused y".

Researchers in other fields have recognized this issue and a summary from one prominent researcher in this area is provided in the following:

> *"Why is causality important? If an evaluator erroneously concludes that a program is meritorious (because it is thought to have caused some positive changes), resources may be wasted on continuing it or expanding it in its current form. ... Conversely, a good program might be discontinued or altered if negative changes are wrongly attributed to it.... In other words, causality is not merely an issue of relevance to academics; it deeply affects the lives of many stakeholder groups, whether they realize it or not."* [c]

The approach that has been taken toward assessing causality to date has been one that looks at the chain of program interventions, expected outputs, and outcomes in order to determine causality. The NYSERDA *September 2000 Interim Evaluation and Status Report* added clarification, that evaluating causality would be assessed from behavioral information that looked at why and how changes in decision-making and policies were made (with implied inquiry as to the program's direct effect and effects from potential intermediate outcomes). This report also stated that the causality assessment will help to determine "*the causal relationship between **New York Energy $mart**[SM] program intervention strategies and sustainable changes within the marketplace"* and "*causality of program interventions both on an individual as well as aggregated (portfolio) level."* [d]

Methodological Approach to NYSERDA's Causality Assessment

Definition and Background Into Causality

The methodology for assessing causality has been constructed based upon professional evaluation experience, best practices, the latest literature in the evaluation field, and the diversity of **New York Energy $mart**[SM] program interventions, including market transformation.[e] The causality approach has been based upon two research examinations: (1) Program Theory

[c] E. Jane Davidson, "Ascertaining Causality in Theory-Based Evaluation," *Program Theory in Evaluation: Challenges and Opportunities*, eds. Patricia Rogers, Timothy Hacsi, Anthony Petrosino, and Tracy Huebner, pg. 17, New Directions for Evaluation, American Evaluation Association, Number 87, Fall 2000.

[d] NYSERDA, *Evaluation and Status Report, Interim Report, September 2000*, pg. 2-14.

[e] Causality within the energy efficiency market transformation (MT) field is still often discussed more on a theoretical basis then with proven methods and tests. This difficulty is compounded for the **New York Energy $martK** evaluation given NYSERDA's lofty goals of assessing the broad causality parameters of behavior in such a large and multi-faceted effort involving a portfolio of energy efficiency services, low-income affordability, and research and development programs.

Evaluation and, (2) Quasi-Experimental Design[f]. The merits of each of these examinations are briefly outlined below.

Program Theory Evaluation. Program Theory Evaluation (PTE), also referred to as Theory-Based Evaluation (TBE), charts the flow of activities from an intervention to an outcome to further outcomes as well as the interactions of outcomes, based on a defined program theory. Measuring each step in the causal chain (and expected changes/reactions) provides information that can separate problems with the theory of causal effects (the basis of program design) from program failure to set a stage in motion. This is best illustrated in Figure 1.[g]

Figure 1, Program Theory Failure Versus Program Implementation Failure

Deleted: :

Program logic models can assist in the causality assessment as the predicted path of occurrences from program intervention as each examined more closely. The use of PTE for this purpose is best described as follows:

> *Developing a clear picture of what has occurred becomes an essential part of developing an understanding of why it occurred (i.e., assessing causality). One of the best methods of examining causality is by identifying the linkage between possible or expected consequences (outcomes) from each program intervention (program inputs and activities). Program theory and logic models provide the basis for this linkage. ...Movement along a continuum of expected outcomes can be examined to determine how market actors made decisions and how these decisions might have affected future decision actions. Viewing decisions in this way, allows both causality and sustainability to be at least partially assessed."*[h]

Program Theory Evaluation (PTE) and the use of program logic models support the belief that detailed models, such as those that include socio-behavioral processes, can provide evidence of causality w hen a change occurs directly where there has been an intervention. While this assists in the assessment of causality, it does not prove causality as thoroughly as an experimental or quasi-experimental design could. This is because as the program intervention is examined and outcomes measured, it is difficult to provide evidence that the program caused the change (created the causal process) rather than any other factor (e.g., continuing previous trends, increased energy prices, news on California energy crisis, etc.). Having an outside factor entirely cause the outcome observed is another type of program failure as shown in Figure 2. In this circumstance, the failure to recognize that the cause was actually a third-party outside occurrence lessens the internal validity to concluding program effects caused the change.

[f] See Thomas D. Cook and Donald T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Houghton Mifflin Company: Boston.

[g] Weiss, Carol H., 1997. "How Can Theory-Based Evaluation Make Greater Headway?" *Evaluation Review*, Vol. 21, No. 4, August, pp. 129.

[h] NYSERDA, *Evaluation and Status Report, Interim Report, September 2000*, pg. 2-15.

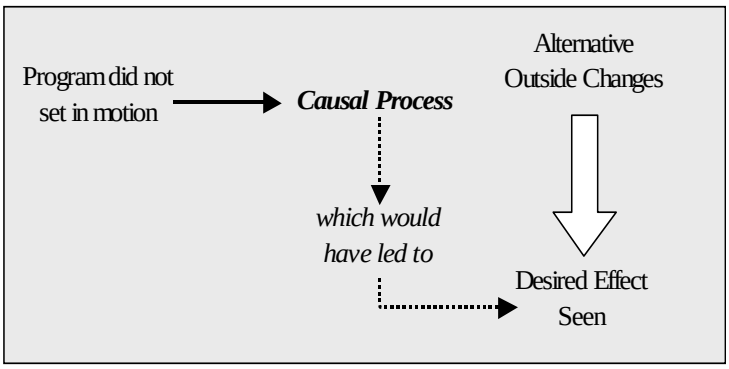

Figure 2. Desired Effect Observed, But Program Did Not Cause Change

Quasi-Experimental Design. Quasi-experiments refer to experiments that have treatments, outcome measures, and experimental units but do not use random assignment to create comparison groups. Then other research design elements are used to assess the treatment impact and assess threats to valid causal inference. There are a variety of research techniques used to assess treatment impacts (the hypothesized causal factor of interest) from other potential causes, and to do so by adjusting for differences in the comparison groups. This is the basis for many statistical techniques (e.g., regression analysis, Analysis of Variance, etc.).

Most of the market transformation activities within the **New York Energy $mart**[SM] effort have program measurement that include a baseline measurement and then one or more market progress measurements. In the style of a program design diagram, this simple time line view is presented in Figure 3.

Figure 3. Basic New York Energy $mart[SM] **Baseline-Market Progress Design**

This simple pre-post design (with continuing treatment) does not address the possibility of other causes in changes between the baseline and post-intervention periods (differences between O_1 and O_2, O_3, etc.). A comparison group, preferably by random assignment, would be one way to do this. Yet, random assignment comparison groups can seldom be used in public policy.

Combining Program Theory Evaluation and Quasi-Experimental Design. The **New York Energy $mart**[SM] causality assessment is mindful of the benefits of examining the causal process embedded within the program logic model(s) and ensuring that outside factors, if they exist, are not the only factors creating the changes being seen. This requires a mix of Program Theory Evaluation (PTE) techniques and quasi-experimental design market comparisons. This hybrid approach involves gathering evidence in support of causality from a variety of direct and secondary research efforts.

Viewing program effects both internal and external to a program effort pushes the assessment to be one focused upon markets. A market-based causality assessment needs to simultaneously consider a broader portfolio of **New York Energy $mart**[SM] efforts that might affect that market. All programmatic influences towards the desired outcome in that market should be considered.

There are also other possible factors that must somehow be considered to ensure the evaluation does not attribute the program with an effect that is actually caused by another (external to the program) change in the market. There are generally three approaches that can be used to include consideration of outside factors. These include:

- Directly asking market participants if it was the program that caused their actions (as opposed to other factors);
- Identifying the likely outside factors and measuring their possible effect on the desired outcome; and/or;
- Conducting the assessment as an experimental or quasi-experimental evaluation. This can include comparisons in the desired outcomes across similar markets having different levels or timing of program intervention.

The information to be used in performing NYSERDA's causality assessments will be derived by: (1) adding desired information to the data collection activities already occurring in NYSERDA's program-specific market and program measurement efforts; (2) conducting market studies and additional research efforts; and (3) gathering secondary data from other regions or nationally as comparison data. The causality assessments will then be a synthesis obtained from harvesting and comparing information from the market progress efforts, direct causality questions posed to mid-stream and other key market actors, intermediate outcome indicators, and regional result comparisons.

NYSERDA's Current Causality Examination Approach

The current approach taken by NYSERDA to determine causality involves all three levels of information sources as discussed above. While the concept of causality is theoretically grounded, the factual determination of programmatic influence with a high degree of certainty has not been widely adopted or proven on a wide scale for energy efficiency program investments. Because this approach is just now beginning to be used as an evaluation activity by NYSERDA, it is expected that there will need to be regular clarifications of the causal process, as well as continual feedback at each step of the process, so that the causal evaluation may be able to react to unforeseeable data needs or external events. A four-step process has been developed and is currently being employed to begin evaluating causality for the **New York Energy $martK** program. This process is outlined in Figure 4. These steps include:

Step 1: Enhance Ongoing Data Collection Activities.

- Identify the specific surveys, research and program-specific evaluation activities being performed by program implementation contractors and discuss modifications/enhancements to help with causality assessment.

Step 2: Identify Additional New York-Specific Market Studies and Research Efforts.

- Identify ongoing & planned market studies and additional research efforts to be undertaken for the purpose of gathering NY-specific data to help assess causality.

Step 3: Identify Regional and National Comparative Information.

- Discuss plans for collecting regional & national comparative information.

Step 4: Data Verification and Analysis and Reporting.

- Assure data collection was consistent with industry standards and that the analysis was objective and non-biased by external events.
- Report results to programs, external stakeholders, and customers.
- Utilize this process as a feedback mechanism to continually improve the causality assessment.

Figure 4. Current Causality Approach at NYSERDA

These four steps are expected to lead to a mixed assessment of causality by enhancing and adding to current measurement and evaluation efforts. This will combine information sources from directly within the program theory logic models for **New York Energy $martK** program efforts, program participant inquiries, external comparisons, and other quasi-experimental design elements as provided by program activities. It will create and gather the three levels of information described in the prior section.

An example of how these different elements are being used is provided in Figure 5, which shows the way in which causality is being examined in the ENERGY STAR® appliance and lighting markets. A portion of the program logic model is displayed, as well as outputs examined along the path diagram (awareness, retailer changes, and market share), and the causal determinate examinations being made through quasi-experimental design comparisons. Each of these is further described below within the presentation of the early evidence of causality regarding the **New York Energy $martK** efforts in the residential appliance and lighting markets.

Figure 5. PTE and Quasi-Experimental Examinations within the Causality Assessment of the ENERGY STAR® Appliance and Lighting Markets

Causality Evidence from the Residential Appliance and Lighting Market In New York

Consumer Awareness

The Consortium for Energy Efficiency (CEE) sponsored a national household survey on ENERGY STAR® label awareness, understanding, and influence during the summer of 2000. Designed Marketing Areas (DMAs) were classified as being "high message saturation", "low message saturation", or in-between (other). The **New York Energy $martK** program area overall would be classified a high message saturation area. The CEE study found that label awareness was "much higher in the high-publicity areas (such as New York) than in the low-publicity areas - 52 percent versus 37 percent (p-value <0.0005)."[i]

The **New York Energy $martK** Residential Appliances and Lighting program, through implementation contractor Aspen Systems Corp., conducted baseline and market progress measurement in 1999, 2000, and 2001. A household mail survey was a large part of this effort. The

[i] Goldberg, Miriam, Mitchel Rosenberg, Marc Hoffman, Tim Pettit, and Maureen McNamara. 2001. "Counting the Stars in America's Eyes: The ENERGY STAR® Household Survey," *Proceedings of the 2001 Energy Program Evaluation Conference*: Salt Lake City, UT, pp. 350.

awareness questions in these surveys were significantly different than in the 2000 CEE survey. Yet, the results are quite similar. Awareness for New York state, according to this repeated survey, has increased from 34% in 1999 to over 43% in 2001.[j]

An experimental design for causality would be in starting the intervention at a different time or at different levels in different geographies. This was not done, but by happenstance, different areas in New York did receive slightly different treatments. Differences in outcomes in the expected directions given different treatments are, therefore, strong quasi-experimental evidence of causality of the treatment.

The Residential Appliances and Lighting program's 2001 report several times mentions the circumstances that led to the lower awareness gains seen in the data for downstate New York. These include the fact that NYSERDA did not allocate budget for downstate advertising by DDB Worldwide Communications (the contractor performing the ENERGY STAR® Public Awareness Campaign). This is because, at the beginning of **New York Energy $martK** effort, it was understood that the U.S. Environmental Protection Agency (EPA) planned to produce advertising for downstate New York, and advertising in this market is quite expensive. These facts made it prudent for the program not to allocate program dollars for downstate. EPA did not end up advertising downstate. This was unfortunate for the program, but did provide a subsequent quasi-experimental design for causality. The quasi-experimental design diagrams for these comparisons are shown in Figure 6.

Figure 6. Quasi-Experimental Design within New York ENERGY STAR® Appliance and Lighting Markets

ENERGY STAR® awareness in the four upstate DMAs rose from 34% in 1999 to 54% in 2001. During the same time period, the area without the completed media campaign through New York's ENERGY STAR® Public Awareness Campaign, downstate New York, had ENERGY STAR® awareness only grow from 34.5% to 37.2%.[k]

Looking at the Buffalo DMA provides a second quasi-experimental design regarding awareness. "Buffalo was the first DMA to receive the Program's ENERGY STAR® marketing efforts. Buffalo has

[j] Final Project Report, *New York State ENERGY STAR® Appliances and Lighting Program, Phase II, Task 6*, Prepared for New York State Energy Research & Development Authority by Aspen Systems Corporation, August 9, 2001.

[k] Ibid, pp. 2.6 - 2.7.

also had exposure to the logo through several special events that were not replicated to the same degree in other DMAs." Awareness in the Buffalo DMA grew the most from 1999 to 2001, and to the highest level in the New York study, 56%.[l]

Further corroboration of the national findings and for New York's causality assessment is provided by the fact that the lower level of media activity downstate resulted in 37% awareness, and maximum activity and length of activity (in the Buffalo DMA) resulted in 56% awareness. These figures are remarkably close to the 2000 CEE study with 37% awareness in low publicity areas and 52% in high publicity areas.

The 2001 Aspen study also reported responses for 2000 and 2001 to the questions of "Have you ever seen any advertising featuring the ENERGY STAR® logo?". This data also shows the same relative findings and trends. The Buffalo DMA increased the most and to the highest level rising from 17.8% in 2000 to 39.8% in 2001. New York City went from 12.4% to 16.9%. The state overall was in between the two, with 15.1% in 2000 and 23.9% in 2001 seeing advertising with the ENERGY STAR® logo.[m]

Another indicator of the validity of the quasi-experimental comparisons made here can be seen by detailed examinations of where the New York household survey respondents say they saw the ENERGY STAR® logo. This was done by comparing ratios for three areas: upstate New York, the non-DMAs, and downstate New York. The working hypothesis given the quasi-experimental situation described above, is that the ratio examinations should show that media advertisements are increasingly more important among those aware of the ENERGY STAR® logo in areas with greater media activity. This means the ratios should point to media advertising being more important in the following order: downstate New York, non-DMAs, followed lastly (most important) by upstate New York. This factor should also increase over the 1999 to 2001 time period as the greater media activities are employed.

This, in fact, has occurred. In 1999, the point spread across the three areas in printed material to number aware was four to six percentage points. As 1999 did not separate media advertising from print media, the 2001 examination looks at two proportions: printed material to media advertisement, and media advertisements to number aware of logo. Comparisons across all three of these proportions are presented in Table 1. All support the quasi-experimental hypothesis, showing that where there is more ENERGY STAR® media activity there is a greater effect.

[l] Ibid, pp. 2.4 - 2.5.
[m] Ibid, pp. 2.22.

Table 1. Causality Evidence from Comparison of Media Advertisement Effects*

	1999 % of those aware of ES logo that learned it from printed material	2001 Proportion learned from printed material to proportion learned from media ads	2001 % of those aware of logo that learned it from media ads
Downstate NY	2.2%	31%	16%
Non-DMA	8.2%	28%	18%
Upstate NY	6.7%	11%	25%

* Data from 1999 Baseline Report, and Final Project Report, *New York State ENERGY STAR® Appliances and Lighting Program, Phase II, Task 6*, Prepared for New York State Energy Research & Development Authority by Aspen Systems Corporation, August 9, 2001.

Retailer Reports

"Step 1", enhancing already on-going measurement efforts, is an important part of NYSERDA's causality approach and ensures cost-effective evaluation. The 2001 Aspen surveys were being refined as the causality approach was being developed. The opportunity was used revising the survey to include a couple of questions directly asking retailers whether they would have undertaken the ENERGY STAR® activities even if NYSERDA had not been involved (the program counter-factual, what would have occurred if the program had not happened). The results from this inquiry are presented in Table 2.

Table 2. Direct Retailer Responses on Counter-Factual ENERGY STAR® Activities*

	Appliance Retailers	Lighting Retailers	Home Electronics Retailers
Definitely would have increased ENERGY STAR® stocking & promotion practices without NYSERDA	18.6%	12.5%	33.3%
Might or might not have increased them**	27.9%	50.0%	30.0%
Definitely would not have increased them	48.8%	37.5%	30.0%
Did not respond or Don't Know***	4.7%	0%	6.6%

* Data from 1999 Baseline Report, and Final Project Report, *New York State ENERGY STAR® Appliances and Lighting Program, Phase II, Task 6*, Prepared for New York State Energy Research & Development Authority by Aspen Systems Corporation, August 9, 2001.

** Often interpreted as a probability around 40%, given not knowing meant it was not already planned or budgeted.

*** Often interpreted as not likely to have done so.

Market Share

The 2001 *New York State ENERGY STAR® Appliances and Lighting Program, Phase II* report included a time trend comparison that provides evidence of causality with regard to market share changes. This was done on the four appliances where data was available from the Association of Home Appliance Manufacturers (AHAM): refrigerators, dishwashers, clothes washers, and room air-conditioners. A study of market share growth from the Baseline Report in 1999 through 2001 was compared to a forecast of market share in 2001 given AHAM trends. With all four appliances, the New York market share growth was significantly greater than the AHAM forecast.[n]

Conclusions for Residential Appliance and Lighting Market Causality Assessment

There are at least five separate examinations in this assessment that provide evidence that the **New York Energy $martK** is causing the outcomes being reported from the program theory, the increase in awareness, retailer activity, and market share in the residential lighting and appliances markets. This is the most extensive evidence of causality the energy efficiency field has seen to-date. It also supports the causality assessment approach being used by NYSERDA as a feasible way in which to conduct its causality assessments.

References

1. Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Houghton Mifflin Company: Boston.

2. Davidson, E. Jane. "Ascertaining Causality in Theory-Based Evaluation," *Program Theory in Evaluation: Challenges and Opportunities*, eds. Patricia Rogers, Timothy Hacsi, Anthony Petrosino, and Tracy Huebner, New Directions for Evaluation, American Evaluation Association, Number 87, Fall 2000.

3. Goldberg, Miriam, Mitchel Rosenberg, Marc Hoffman, Tim Pettit, and Maureen McNamara. 2001. "Counting the Stars in America's Eyes: The ENERGY STAR® Household Survey," *Proceedings of the 2001 Energy Program Evaluation Conference*: Salt Lake City, UT.

4. New York State Energy Research & Development Authority (NYSERDA), *Evaluation and Status Report, Interim Report, September 2000.*

5. New York State Energy Research & Development Authority (NYSERDA), Final Project Report, *New York State ENERGY STAR® Appliances and Lighting Program, Phase II, Task 6*, Prepared for New York State Energy Research & Development Authority by Aspen Systems Corporation, August 9, 2001.

6. Weiss, Carol H. 1998. *Evaluation: Methods for Studying Program and Policies*, Prentice Hall, Upper Saddle River: New Jersey, Glossary.

[n] Ibid, pp. 2.60.

7. Weiss, Carol H. 1997. "How Can Theory-Based Evaluation Make Greater Headway?" *Evaluation Review*, Vol. 21, No. 4, August.

Address any questions or additional copies of this paper to:

Dr. Lori Megdal
Megdal & Associates
198 High Street
Acton, MA 01720-4218

Telephone: 978/461-3978
Fax: 978/461-3979
E-Mail: megdal@bellatlantic.net