"Theory-based Evaluation for Market Transformation"

Dan Violette

IL EE Stakeholder Ad*visory Group: Market Transformation Savings Working Group Small Group*

Wednesday, October 26, 2022
9:00 – 10:30 am

The Context of MT Evaluation

- *Quote used in: Jim Fay, ComEd; Rick Tonielli, ComEd; Randy Opdyke, NiCor Gas; Patricia C. Plympton, Guidehouse. "***Making Market Transformation Mainstream in Illinois**." Proceedings of 2022 Summer Study—the 22nd biennial ACEEE Summer Study:

  "Existing program designs have also been heavily influenced by evaluation, measurement, and verification (EM&V) beliefs and practices that less and less reflect energy efficiency policy aims, advances in data analytics, or the growing understanding of customer behavior. This has created ***a bias toward programs that are easily evaluated****, disincentivizing utilities to explore more innovative program designs that would require complicated analysis to determine program performance."*
  Val Jensen, Senior Fellow, ICF (July 2022)

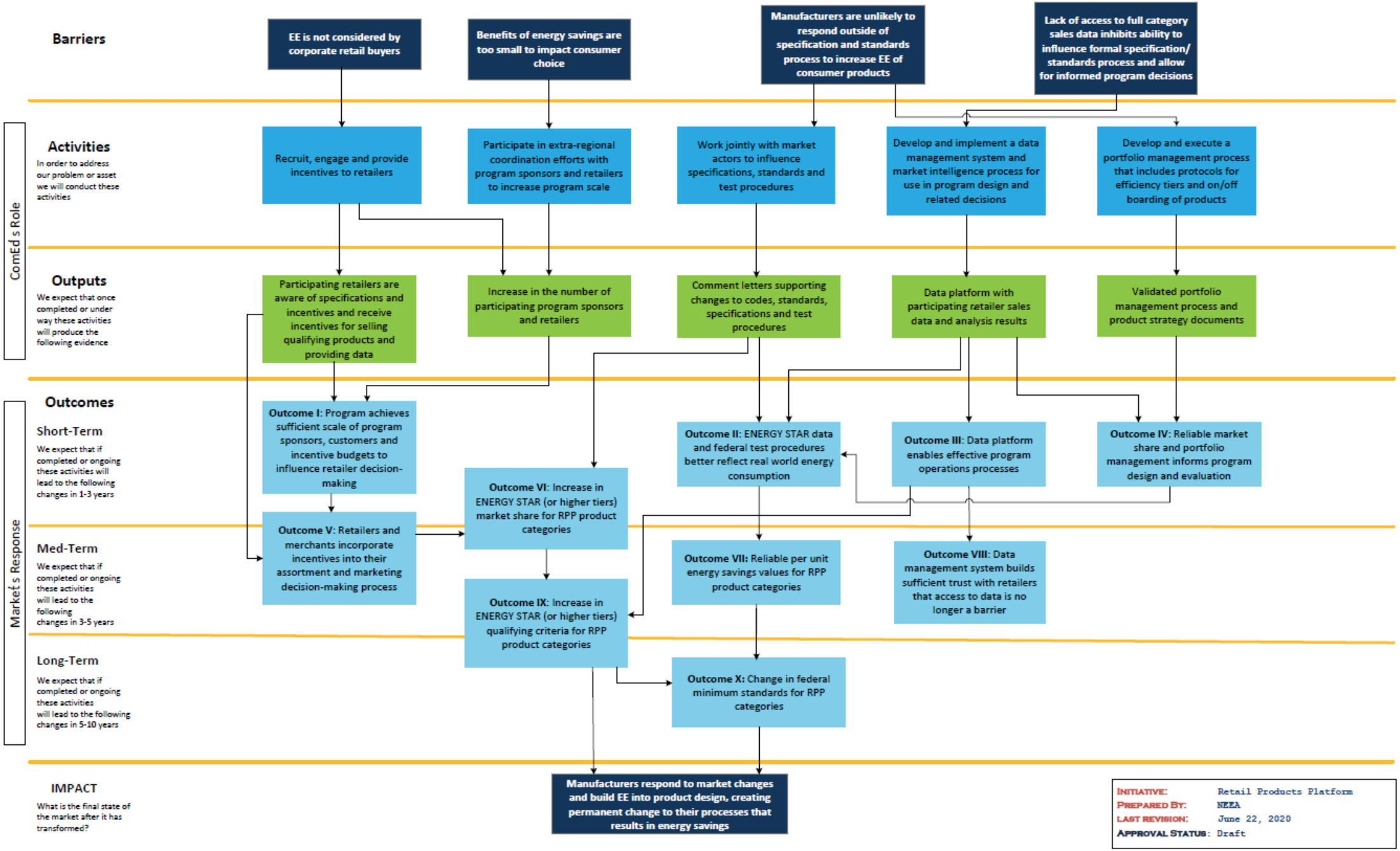
- Early DSM efforts often lent themselves to specific evaluation frameworks:
  - Commercial lighting where ΔWatts x Hours can be measured.
  - Smart/programmable thermostats with thousands of participants and non-participants.
- Meeting aggressive clean energy goals may require more complex program designs that don't lend themselves to annual sampling of participants and non-participants.
- However, many important investment and resource allocation decisions across industry also fall into this category.

INTRODUCTION: Theory-Based Approaches to Evaluation

- Theory-based evaluation is an approach and not a specific method or technique. It is a way of structuring and undertaking analyses in an evaluation.
- Theories of change illustrates "the story of what should happen in the 'arrows' that link the boxes in a traditional logic model."[1]
- Theory-based evaluation develops "the assumptions, risks and context that support or limit the theory from being manifested as observed outcomes."[1]
- The theory of change can be used to test — with evidence — the assumed causal chain of results with what is observed to have happened, checking links and assumption in the process to verify the Logic model.

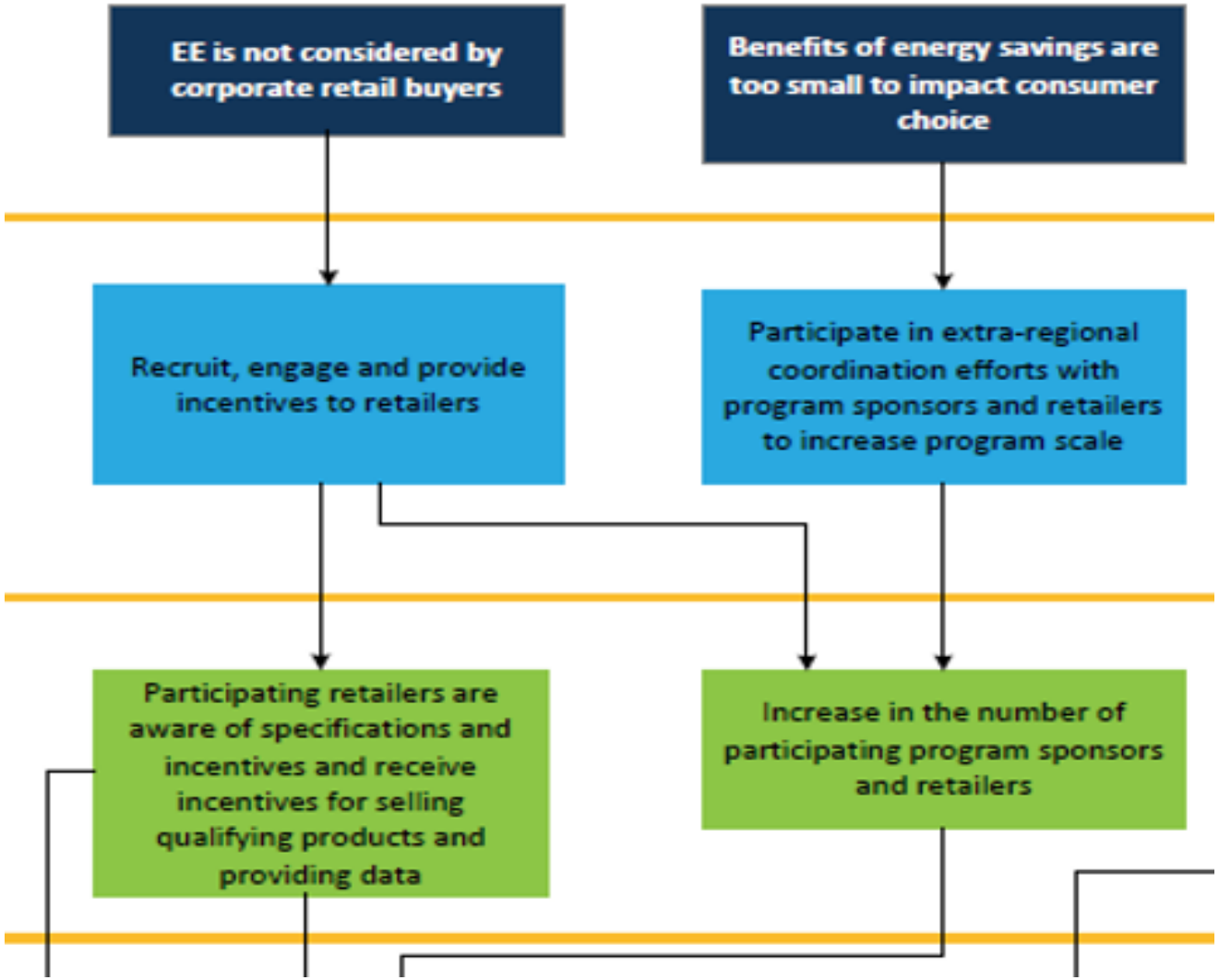
[1/] Theory-Based Approaches to Evaluation: Concepts and Practices. Treasury Board of Canada. https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/theory-based-approaches-evaluation-concepts-practices.html

ComEd ENERGY STAR® RETAIL PRODUCTS PLATFORM LOGIC MODEL



INITIATIVE:	Retail Products Platform
PREPARED BY:	NEEA
LAST REVISION:	June 22, 2020
APPROVAL STATUS:	Draft

A Subset of the Logic Model – Validating Logic



- NYERDA used Core and Key Indicators in a hierarchy used in the Market Characterization & Assessment (MC&A) evaluations.
- These indicators included availability and awareness across technologies and applications.
- One factor examined was confidence that the measures would work, i.e., deliver the promised change in energy use.
- "**Non-quality**" in delivery and market activities has been found to be a substantive leak in cost-effective energy savings.
- Tracking Market Progress indicators is at the core of any Theory-Based Evaluation.
- But is that enough?

Burden of Proof

- What does a “preponderance of the evidence” mean?
  - One suggestion is that it means a greater than 50% probability of hypothesis being true, i.e., a fact/hypothesis is more likely to be true than false.
  - What hypothesis is being tested – how do you structure the question?
    - Impacts of “X“ kWh was attained, i.e., is it an energy savings number or is it the baseline?
    - Can influence be detected, i.e., are the expected trends over time being observed.
    - Is the program cost-effective, i.e., what level of impacts and influence are needed to achieve cost-effectiveness.
- The preponderance of the evidence approach requiring that a claim is just more likely to be true than false seems a little unsatisfying – shouldn’t we ask for a little more?
- Consider preponderance in the legal context:
  - Often used in civil cases, but how is this determined, i.e., who decides?
    1. A specific claim is developed.
    2. A jury is selected (often 6 or 12 people).
    3. A super majority of the jury must agree that the hypothesis is more likely to be true than false. Usually, 9 or 10 out of 12 jurors but, for example, it must be unanimous in Illinois.

Burden of Proof (cont.)

- Does this view of "preponderance" provide a little more confidence than just the simple statement that "it must be more likely to be true and false."
  - Legal comparisons are always a bit dangerous, but is there an analogy?
  - Does having a supra-majority of "reasonable" people accept the preponderance criteria add a layer of confidence?
- Major resource allocation decisions are made throughout our economy using judgement and expert opinion – not a random control trial (e.g., pharmaceutical-grade evidence).
- Like the supra-majority above, what can increase our confidence:
  - Regular tracking with mid-course corrections – you don't roll a program out and then ignore it – this is important to recognize as it helps manages risk.
  - Does a good Logic model supported by Market Performance Indicators provide the needed confidence.
  - What about small sample verifications – follow 10 to 12 installations and validate the savings in the field? (Note: NYSERDA employed this approach)
    - These field verifications can sometimes find unexpected threats.
    - Simply looking at installations can be important.
  - If we have a structured expert panel, does that help?

Theory-Based Evaluation compared to current approaches?

- Aspects of attribution estimation are currently driven by the logic model.
  - Looking at Attachment A: Illinois Statewide Net-to-Gross Methodologies, the term “logic” appears 20+ times, and usually supports judgmental decisions on how to develop program influence scores.
  - In the current “Illinois Statewide Net-to-Gross Methodologies” you see statements such as:

    “If these sources of participant spillover that are not captured are expected to be large (based on the best research available or given the program’s logic model), adjustments or additional analysis to capture these types of participant spillover may be required.”
  - These approaches use structured surveys of market actors.
  - A View: Given the use of logic models and the corresponding program theory in the evaluation of attribution for RA programs, could this imply:
    - That these approaches are not new, and
    - The use of these methods in MT evaluation is more of an extension or a different emphasis rather then the application of new methods.

The Role of Structured Expert Judgment Approach

- There are lots of business and resource allocation problems across industry that do not lend themselves to experimental design driven assessment.
- Expert judgements and structured decision making are often used.
  - These approaches are used in finance, R&D for drug development, environmental policy, business strategy.
  - Companies work with their senior executives to assess what they believe about future markets and strategies, and develop strategies consistent with these beliefs.
  - It to construct range and uncertainty estimates from industry experts that can be used to develop confidence intervals.
  - It can integrate information from multiple sources.
- There are examples in energy savings evaluation literature.
  - The DOE UMP NTG chapter had a couple of examples.
  - However, there are not as many applications as might be expected.

Approaches and Context for Use of Expert Judgment

- One interesting application involved the evaluation of the Wind Power America program.
  - The goal was to assess the impacts attributable to the program.
  - The number of wind farms built did not provide for participant and non-participant groups that could form an experimental design – but estimates were needed for assessing the viability of the program.
  - One aspect of this study was the use of range estimates:
    - Experts were asked about lower and upper bounds around the program effects; and, then a best estimate. This approach allowed the experts to provide insights into amount of uncertainty of the estimates.
    - Gauging uncertainty and then using that information in probabilistic and scenario analyses are consistent with other utility resource planning activities.
    - Adapting these methods to EE resource assessment may increase the usefulness of evaluation results.
- CONTEXT: Supply-side resource decisions at utilities have long required expert judgment on important inputs such as fuel costs, carbon adders, risk factors (e.g., reliability), costs of construction, operational factors, etc.

Basic Approach to Range Estimation

- Provide background and define the problem to be addressed.
- At the beginning, ask experts to consider factors likely to affect energy savings and to what degree (provide list of factors known to affect energy savings).

  0 – little to no impact

  1 – some impact

  2 – significant impact

  - This allows the experts to think about what might cause the true value of the energy savings to vary up and down – one bias avoidance approach.
  - Information may be provided to experts from program records.
- Start by bounding the range:

  1) Ask what the likely lower bound might be for the actual energy savings value. Repeat for upper bound.
  2) Likelihood estimate - split the energy savings range into three bins and ask the interviewee to rank these likelihood of the true value falling into one of these bins.

Structured Expert Judgement

- Structured expert judgement approaches have not had the success in the EE/DSM Evaluation field. Why?
- It may be the case that the experience with these methods has not been what was needed – protocols for these methods have been lacking in our field.
  - People in our field focus on their specific areas of expertise in engineering or statistics.
- It may reflect the history of evaluation and the expectation for experimental design driven approaches.
- There are a few examples where EE/DSM decision makers that used the results of these studies:
  - NEEA Sunset Study
  - Iowa Utilities Study
  - Ontario Study

Structured Expert Judgement – Some Success Factors

- Thoughtful selection of experts – much revolves around how you organize the results if you get info on the underlying factors that influenced their responses.
- Do not ask the experts questions that they simply can not answer, e.g., don't require too much resolution around the distribution of energy savings impacts.
- Use review and feedback to help ensure that an appropriate range of factors are considered.
- Use range estimates for impacts (sometimes referred to as "fuzzy" Delphi).
- Don't ignore the expertise of the program administrators as they can provide useful input:
  - It may not be appropriate to include program personnel responses with the expert findings.
  - But, by having program personnel participate in the surveys, gaps in the analysis can be identified and used to improve future evaluation efforts.
- Get the experts to provide information on the factors that produce the lower bound and upper bound estimates – what drove their estimates of the range of impacts.
- At the end, ask the experts if the process was credible and represents their views.

Conclusions

- We shouldn't let available evaluation approaches overly influence our decision to implement those program concepts that are likely to be most cost-effective.
- There is a need to develop an appropriate context around MT savings estimates – we can use best practices to produce information that supports good decision-making regarding EE/DSM investments.
- Market Progress Indicators (MPIs) likely will be the core of a theory-based MT evaluation; but, it may not be enough.
- Consider augmenting MPIs with:
  - Small-sample infield studies with 10 to 15 sites – high validity sites.
  - Structured expert judgement approaches.
    - If done correctly, the expert judgement process will be informative in identifying those factors that create risk for attained energy savings.
    - It allows for the bracketing of uncertainty in energy savings estimates.

Contact Information

Dan Violette

Director and Senior Advisor, Apex Analytics, LLC

danv@apexanalyticsllc.com

Phone: 303-903-3035 (cell)