# Evaluation in a Messy and Changing Environment: Maintaining Focus on the Evidence-Based Role

*Ingo Bensch, Resource Innovations, Madison, WI*

## ABSTRACT

This paper is intended to be a thought piece on (1) how the evolution and rapid changes in the energy industry affect the practice of evaluation and (2) how the energy evaluation field may respond to continue to contribute to sound programs, initiatives, and policy.

The paper presents observations of key ways that the energy field has become more complex for programs and evaluators. Energy programs and initiatives are being called upon to meet increasingly broader sets of policy objectives, ask consumers to engage with energy in more complicated ways, and they are doing so in a more complex social environment.

Navigating these complications is a challenge not just for energy programs, but also for evaluators who are being asked to answer new questions and help guide programs through more complex sets of objectives. The paper offers thoughts on ways the energy evaluation field is already well-suited to respond to these challenges and ways it needs to adjust. Ideas offered include the development of new tools and more sophisticated coverage of policy objectives when evaluating programs.

The paper closes by suggesting that these considerations should be addressed collaboratively by the energy evaluation field and offers some steps for doing so.

### Introduction

This paper is a thought piece intended to spur reflection and discussion by evaluators of energy efficiency programs and similar ratepayer and publicly funded initiatives. These energy programs and the context in which they operate keep evolving. Many of the changes the energy efficiency field has experienced in recent years have introduced greater complexity, which leads to new issues for evaluators, policymakers, regulators, program implementers, and advocates.

Consider, for example, that the same energy efficiency rebate programs that were once created and funded to delay the need for new power plants under the initial framework for demand-side management (DSM) programs are now justified by broader considerations. These considerations still include overall cost-effectiveness for ratepayers in both regulated and non-regulated utility environments. They also now include a host of newer policy considerations and related industry developments such as climate-related objectives and goals, the integration of distributed energy resources, electrification of buildings and transportation, load flexibility, resilience, workforce development, and equity (while striving to maintain broader energy affordability).

Arguably, energy evaluation practices have a strong conceptual framework for assessing overall costs and benefits for energy supply and energy demand management, while the shift to addressing newer policy considerations is a work in progress. Readers may have their own perspectives on these topics. The point is that there has been change and a need to adjust to the new demands on evaluation and opportunities for our field to contribute to sound policy, programs, and market interventions.

In addition, the policy context behind energy evaluations and programs is increasingly framed in political contexts with elected officials, influencers, and other thought leaders pushing their audiences to think of energy, climate, market interventions, and equity in increasingly political, value-laden, or divisive terms. This dynamic is a challenge for formative evaluation as well as for evaluators themselves as the topics being studied touch increasingly on personal values in more direct ways than cost-effective calculations, thereby introducing a more human element to an empirical practice.

This paper seeks to explore these dynamics more fully and consider what they mean for the practice of effective and useful evaluation. The author offers thoughts and perspectives in the hope of spurring introspection and discussion. In the words of American author and observer of organizational behavior Margaret Wheatley, "without reflection, we go blindly on our way, creating more unintended consequences and failing to achieve anything useful." [1] At a minimum, the purpose of evaluation must be to ensure that the programs studied are achieving something useful and avoiding unintended consequences. Evaluators would be wise to heed the advice to reflect rather than go blindly on our way as the energy industry is transforming and the context in which energy programs operate are changing.

### Paper Organization

This paper contains four main sections:

- *What is evaluation anyway?* seeks to provide a common base from which we can think about evaluation in the energy field by answering briefly what evaluation is and describing some of the roles evaluation plays. This section is not meant to present new concepts as much as to ensure that discussion about a mindful evolution of the energy evaluation practice is based on a common base of understanding and terminology.
- *What should we be reflecting about?* lays out thoughts about the transformations in the energy industry and changes in the context in which energy programs operate that warrant consideration by evaluators as we reflect upon the evolution of our practice.
- *Considerations for a mindful evolution* seeks to contribute thoughts about ways the energy evaluation field might respond to the need to continue to evolve (or sometimes return to the basics). This discussion includes consideration of the implications of the evolution we are experiencing and whether or how evaluation can simultaneously respond to the various evaluative needs and opportunities.
- The closing section titled *Next steps* encourages the energy efficiency field to engage in collaborative reflection so that we can approach evolution in the energy evaluation practice thoughtfully and with intentionality.

## What Is Evaluation Anyway?

Most evaluators would probably agree that their practice involves an **evidence-based inquiry** that seeks to document and enhance the effectiveness of programs, initiatives, or interventions (let's call this "programs"). To do this, evaluators examine whether programs are achieving their stated objectives or how they could do so. The evidence-based inquiry is the tool. The program objectives and intervention approaches are the context. Insights concerning how well the evaluated program is achieving its goals or how it could do so is the product.

At the same time, evaluation is not a standard off-the-shelf product. The nature of an evaluation-based inquiry is customized to the program and to the intended use of the evaluation by the entity commissioning the investigation. As a result, there are different types of evaluation studies that serve different purposes. The types of evaluation-related studies are framed somewhat differently across the broader evaluation field. The remainder of this section describes the way that various parts of the broader evaluation field describe the practice.

[1] Retrieved on June 20, 2025 from https://www.azquotes.com/author/17954-Margaret_J_Wheatley.

**American Evaluation Association.** In 2014, the evaluation field's largest and most expansive association—the American Evaluation Association (AEA)—commissioned a task force to define and communicate the value of evaluation. Led by one of the founders of modern evaluation practice, Michael Quinn Patton, the task force defined evaluation as "a systematic process to determine merit, worth, value, or significance (AEA 2014)." Further, evaluators associated with the AEA often distinguish between **formative evaluation** that is intended to inform the goals, interventions, strategy, tactics, and other design considerations for programs, and **summative evaluation** that assesses retrospectively what the program interventions have achieved or are achieving (and what can be learned from that).

**Government evaluation.** The United States General Accounting Office (GAO) distinguishes between program evaluation and performance measurement (GAO 2021). **Program evaluation** is "an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency." **Performance measurement** is "the ongoing monitoring and reporting of a program's accomplishments and progress, particularly towards its pre-established goals." The GAO illustrates the difference using the visual in Figure 1. In addition, it is noteworthy that the GAO describes **policy analysis** as a related endeavor that consists of "a systematic process of identifying and comparing potential options for addressing a policy problem based on certain criteria, and choosing the option that best meets the criteria."

| | What drives it | What data it uses | What frequency | What it can tell |
|---|---|---|---|---|
| **Program evaluation** | Theory of program change | Quantitative or qualitative | Discrete | Whether a program is working and why |
| **Performance measurement** | Agency goals | Typically use quantitative data | Ongoing | How well a program is performing |

Figure 1. GAO description of program evaluation and performance measurement. *Source:* GAO 2021.

Efforts to promote evidence-based governance were further promoted by the U.S. Office of Management and Budget (OMB) during the Biden Administration in response to the Foundation for Evidence-Based Policymaking Act of 2018 (OMB 2021). Directives to federal agencies called for learning agendas and annual evaluation plans, and they also painted a broader vision of evidence-based policymaking of which evaluation is a part. As shown in Figure 2, evidence-based governance is not limited to assessing program performance, but it also includes priority setting, planning and implementation, and driving innovation. In other words, evaluation and its evidence-based cousins can play a pervasive role in the life cycle of a program or policy.

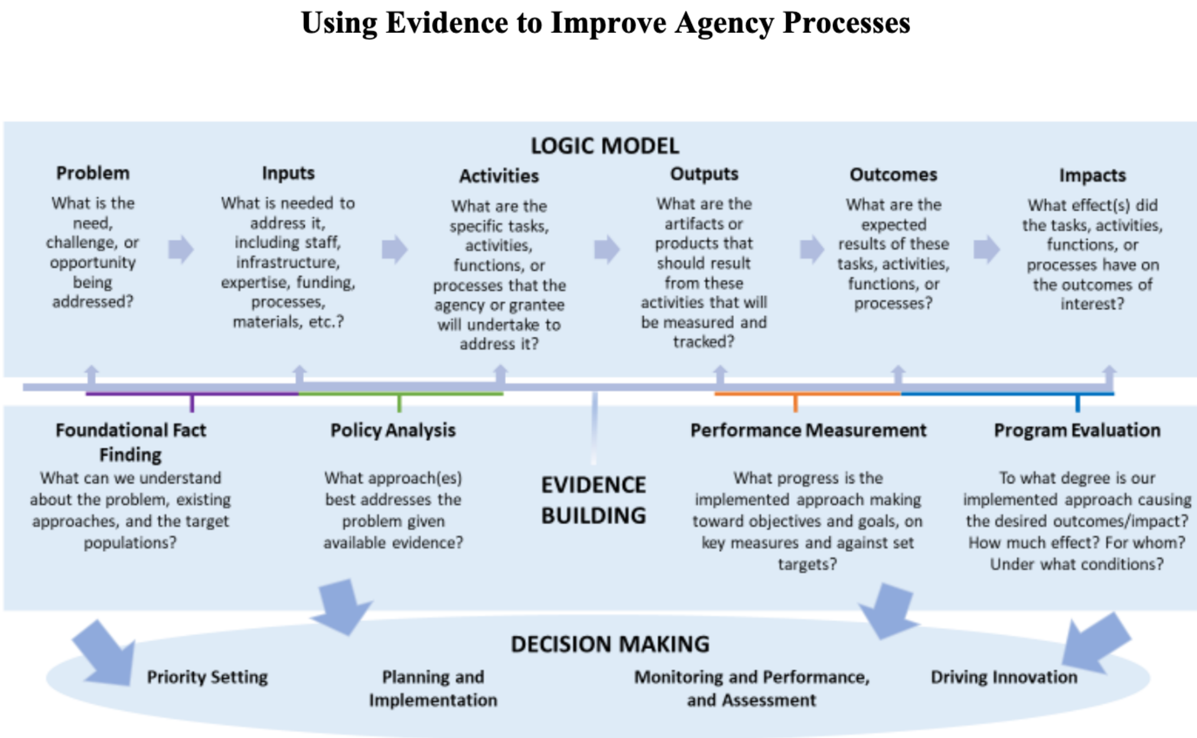


Figure 2. OMB description of how agencies are to use evidence in decision-making. *Source*: OMB 2021.

**Energy evaluation.** The energy evaluation field does not have an authoritative governing body that establishes formal definitions and standards. That role falls to individual public agencies involved in establishing processes and expectations for evaluation work conducted under their oversight or direction as well as various efforts to provide tools and best practices. Generally, common practice in energy evaluation distinguishes between **process evaluation** and **impact evaluation** with **market research** and **technology research** overlapping with these practice areas at times.

Nevertheless, the energy evaluation field has developed some widely used tools specific to impact evaluation. These tools include:

- A series of cost-effectiveness tests for assessing the net benefit of efficiency programs that were first described in the California Standard Practice Manual and have been adopted widely since then (CPUC 2011);
- A general protocol for measuring and verifying savings from energy (and water-) saving projects in buildings and industry called the International Performance Measurement and Verification Protocol (Efficiency Value Organization 2022);
- Jurisdiction-specific technical reference manuals that document approved savings assumptions and calculations to determine energy savings associated with common technology replacements and actions;
- Methodologies used to distinguish free ridership from program-induced impacts;
- A series of 24 protocols provided by the Uniform Methods Project for measuring energy savings from a variety of common efficiency measures (Li 2018).

**Take-aways for considering evaluation's role.** Two themes from this review of existing definitions and types of evaluation seem particularly important during times of evolution and change. They are:

- Evaluation is intended to be an objective, empirical practice.
- Evaluation can serve varying roles and purposes.

The varying roles are important for thinking through the evolution and adaptation of evaluation practice, as the paper discusses in the next section. Figure 3 illustrates some of the roles of evaluation described by AEA, the GAO, and applied in energy evaluation using the terms commonly used in the energy field. The figure describes market research as informing program design, process evaluation as examining the program from design to market interaction, and impact evaluation as concerning itself with the outcomes.

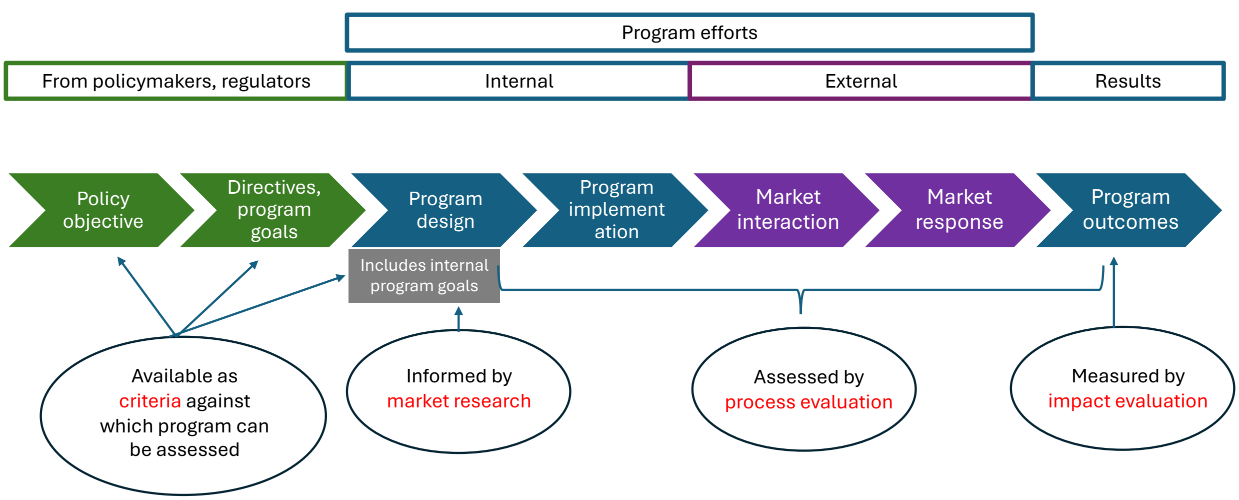


Figure 3. Intersection between program lifecycle and evaluation

The figure also calls out the criteria that are available against which a program can (and should) be measured. It is worth noting that, in energy evaluation, criteria often focus on expectations concerning cost-effectiveness and processes for measuring the net value from using public or ratepayer funds for a program. Linkages of program criteria back to the initial policy objectives are less common and less rigorous and often not addressed directly in individual energy evaluations. As a result, newer policy objectives that have become more central to energy programs' raison d'être are not being reviewed as thoroughly as traditional energy objectives. In contrast, use of policy drivers and directives by policymakers for publicly funded programs is very common in public sector program evaluation, where findings are developed systematically and rigorously against established criteria taken from the programs' policy intent and communicated as such (GAO 2024).

Nonpartisan program evaluators within government tend to be associated with the same bureaus that conduct financial audits. As a result, the standards for developing evaluation findings within governmental program evaluation often are guided by the "Yellow Book" that also guides financial audits. Arguably, rigor in establishing and documenting criteria for program evaluations serves program evaluations conducted within government well. It provides accountability of programs back to the policymakers as well as accountability of policymaker directives and program authorizations against the stated policy intent. Further, by establishing criteria rooted in documented and stated policy intent, evaluators ensure objective, apolitical judgments in a political environment and isolate themselves from having their judgments dismissed because of perceived political bias or leanings.

## What Should We Be Reflecting About?

As noted in the introduction, the context in which energy programs are delivered has changed. There has been an expansion in the policy imperatives programs are charged with supporting, the outcomes they are to deliver, and the basis upon which they are justified. Individual evaluators and stakeholders likely have their own personal perspectives on which ongoing changes matter and how they affect evaluation. The following observations are not intended to be an exhaustive list but rather to spur reflection. Further, the discussion in this section offers a framework for thinking about types of changes that evaluators need to consider.

Recent and anticipated changes in the energy field could be grouped into the following categories:

1. New policy drivers and pre-existing policy objectives that are newly applied to energy programs, resulting in:
    - Changes in individual program goals
        - New and shifting program goals
        - Multiple goals for individual programs
    - Complex interrelationships of policy objectives and programs
        - Policy objectives that rely on multiple programs for success
    - Use of energy programs to achieve non-energy policy goals

2. Technology evolution and changes in their costs that have resulted in a more complex energy system that includes:
    - More varied types of utility generating facilities based on newer and traditional fuels, each of which has a unique set of characteristics for climate, sustainability, reliability, and cost
    - The advent of storage
    - Substantial increases in customer-sited generation
    - Opportunities and complexity in end-uses, as well as the ability to integrate demand and usage
    - Overlap of building energy and transportation energy

3. Changes in values, norms, and public (and personal) sentiment
    - Politicization and culture wars
    - Blurring of the line between research and advocacy
    - Decreasing willingness by the public to engage with legitimate research

Yet, at the same time, some things have not changed much. Programs still rely heavily on rebates or other forms of financial incentives. Arguably, it would be economically more efficient to address the market failures and externalities that voluntary energy programs seek to address with more direct price signals for market actors that truly reflect the societal cost of energy or through standards. However, the use of market forces that reflect true social costs and command and control regulations face political headwinds, resulting in reliance on voluntary programs.

Meanwhile rate designs have become more dynamic, but they are still far less responsive to market factors and societal costs than many other costs that consumers face due to the necessarily monopolistic nature of components of the energy supply infrastructure and the resulting market and regulatory dynamics.

### Changes in Individual Program Goals

As noted, contemporary programs' goals are more varied and more complex. One can think of the shift in program goals as having occurred in three (partially concurrent) stages.

First, traditional demand-side management programs evolved from focusing on deferring needed supply infrastructure investments to pursuing energy efficiency. The energy evaluation field made the needed adjustments to its methodologies successfully. Perhaps most critically for this transition, the prevailing cost-effectiveness tests are well-suited to allow evaluators and regulators to consider varying perspectives and boundaries in which costs and benefits should be considered, so they work well to assess programs in varying contexts and in the pursuit of new savings opportunities as some prior energy-saving measures have become exhausted or obsolete.

Second, energy programs expanded to achieve new goals. Program efforts were added to reduce peak demand, and the energy evaluation field increased the measurement and estimation of demand reductions.[2] Programs emerged to help low-income households afford their energy needs through free and substantially subsidized efficiency upgrades, and energy evaluators modified how benefit-cost assessments are applied to programs serving low-income customers.

Third—and overlapping with the expansion of goals—energy programs' expansive new objectives became layered over pre-existing ones. Efficiency programs became charged with reducing greenhouse gas emissions, decarbonizing buildings, promoting grid resilience through distributed energy resources, and/or promoting a range of narrowly structured and broadly defined equity goals, all the while being asked to support energy affordability. Decarbonization in the interest of climate protection has led to attempts to shift demand structurally and to electrify local combustion of fossil fuels, which is a complex and controversial pursuit that affects a complex infrastructure that was built over many decades.

This is the point where evaluation gets more complicated and in possible need of evolution and adjustment. Some of the goals being added require new ways of thinking about program impacts and consideration of the broader energy portfolio in new ways. In other cases, the individual goals might be measurable, but the optimization among them navigates unchartered waters with little precedent or guidance. Here, new program goals result in a new challenge: dual and sometimes competing objectives. (See also the discussion on dual goals below.)

**Policy goals are being established for which we do not yet have industry accepted metrics and methodologies to gauge effectiveness.** Some goals for energy programs fall outside the domain of energy evaluators' typical toolboxes. Here are a few illustrative examples concerning decarbonization and equity.

Energy evaluators are well-equipped to assess the degree to which a program is supporting decarbonization by quantifying the degree to which non-electrical local loads are being shifted to electricity-powered substitutes. With sufficiently detailed data about time of use and utility generation characteristics during those times, evaluators can quantify the degree to which electrification resulted in actual decarbonization. As such, energy evaluators can answer some core questions about whether and how well programs are supporting decarbonization. However, decarbonization is a complex and future-oriented endeavor. Energy evaluation is not yet well-equipped to address some key questions needed to inform and optimize transitions to carbon-free energy use, such as:

- Is the electrification effort optimized to reduce greenhouse gas emissions when the actual pace of the transition to carbon free electric generation is taken into account? For how long might an early transition to electric end-uses actually be counter-productive?

[2] Interestingly, an increased focus on demand reduction is fully consistent with the original objectives of DSM programs. That renewed focus has been accompanied by technology advancements that better facilitate load management and measurement.

- What is the persistence of bill neutrality promised to customers who electrify if widescale electrification causes rate increases to pay for the required infrastructure upgrades?
- Does the program effectively address trade-offs between goals to reduce energy use and goals to decarbonize?[3]

Separately, policymakers, program administrators, and implementers have committed themselves to equity in their pursuit of energy-related interventions in the market. However, their invocation of equity-related goals is usually broad and aspirational and not sufficiently clear to be actionable or measurable.

Commonly, policymakers and program designers make references—or even set targets—related to program activity among disadvantaged communities, underrepresented communities, historically marginalized communities, environmental justice communities, low-income households, or a mix of these. Fortunately, some of these terms have been defined. California maintains a list of Census tracts that count as disadvantaged communities based on the mix of demographic, health, and environmental factors among residents of those areas; other states (and, at one point, the U.S. government) maintained similar designations using varying terms. And equity has been defined in various publications as encompassing four components: recognitional, procedural, distributional, restorative (Energy Equity Project 2022). While these definitions and classifications are useful, many questions remain unanswered and open to interpretation. Unanswered questions include the following:

Abstract and theoretical constructs

- *Boundaries*: Are equity considerations limited to factors related to income, health, race, and local environmental factors?[4]
- *Resolvability*: If a particular equity concern is invoked as a reason for a program intervention, should the program be expected to create meaningful improvement in that equity concern?
    - For example, if a program targets households in pollution-laden neighborhoods with energy programs, should those programs be expected to reduce the pollution levels that cause the neighborhood to be targeted?

More practically oriented questions

- *Priorities*: Are disadvantaged communities to be considered equal when targeted for energy programs, or are some communities more important to serve than others?
    - For example, do communities with greater energy burden or pervasively low incomes matter more than communities that are designated as disadvantaged due to air pollution or groundwater contamination from industry or agriculture?
- *Transferability to non-residential applications*: How should the concept of disadvantaged communities be applied to business-oriented programs?
    - Is it about the location of the business, the residences of the business owners, the residences of the business employees, the economic footprint of the business (and how that footprint overlaps with disadvantaged communities) or other factors?

[3] This question takes into account the fact that fuel switching itself may or may not save energy when viewed at a site level and needs to overcome transmission losses when seen at a source level.

[4] This question speaks to the fact that equity is a very broad concept but currently applied mostly to a subset of its full meaning, such as to racial equity. Use of the broad term equity imprecisely causes ambiguity about just what is meant and what limits are intended. Without such specificity, the calls to equity could conceivably invoke a much broader concept of fairness in social and transactional interactions of all kinds, including issues related to wealth and power and how interests of people and businesses are balanced.

- *Relationship to actionability*: Should an evaluation addressing the inclusion of underrepresented communities be based on how evenly the program is able to distribute benefits based on pure population counts, or should it be normalized to the degree to which the intervention provided offers potential benefits?
    - And how is "underrepresented" status defined in the first place? Is it based on energy considerations, use of social programs, or other factors?

The main take-away from this exploration is that **new energy program goals and areas of focus lead to new questions that energy evaluators are called upon to address. Some of the narrow tactical program questions can be answered using our field's current approaches; the bigger picture and strategic questions require more thought or input.** (Ideally, well-considered directives would originate from policymakers and regulators, but evaluators need to help point out and sometimes help overcome ambiguities.)

**Dual goals require navigation of trade-offs that we do not know how to optimize.** Optimizing competing objectives and dealing with trade-offs among them will be a challenge whenever there are two or more goals for a program that are not fully coincidental.

This is not a measurement problem, but an optimization issue. When there are multiple objectives, neither one can be maximized. So, the question is not whether the program is as effective as it could be, but rather the challenge for programs and evaluators becomes one of balance and sufficiency. The questions that program designers and evaluators need to grapple with are:

- Is one of the program objectives primary?
- Is there a minimum level of sufficiency that secondary objectives should meet?
- If the goal is to optimize the trade-offs between main objectives, how do policymakers, regulators, and society value the benefit of a marginal unit from one goal against a marginal unit from another goal?

These are not just academic exercises and concepts. The potential trade-offs involved among dual or multiple goals is potentially substantial in some cases. The ACEEE Summer Study paper *Balancing Equity and Energy Objectives: Dealing with the Trade-Offs* sought to demonstrate the potential trade-offs between energy savings, climate benefits, and equity-related metrics when a residential retrofit program seeks to address multiple objectives with a fixed budget (Bensch 2024). The modeling exercise presented in the paper suggested that an equity-maximizing focus could reduce energy and climate impacts by 50 percent in one realistic scenario, and an impact-maximizing focus on cost-effectiveness provides substantially less benefit to prioritized communities than a fully equity-centered approach.

In addition, competing objectives cause ambiguity that program designers grapple with, sometimes with shots in the dark. Without direction and answers to some of these same questions evaluators need to address, program designers do their best to address ambiguous preferences with little precision or strategy.

Consider, for example, a program team instructed to develop a retrofit program that is intended to decrease energy intensity in commercial buildings. That direction alone would prompt the program team to try to maximize energy savings and concentrate on how to best make that happen. Now add policy directives to reduce greenhouse gas emissions and ensure the program focuses on equity. The attention to greenhouse gas emissions introduces electrification measures to the mix of ingredients. Wanting to ensure equity (and lack of clarity on what that means in the commercial sector) may prompt the team to dedicate some of its efforts geographically within disadvantaged communities or expand its focus to include affordable multifamily housing. Adding these ingredients are good faith efforts to address policy concerns, but they may not lead to an initially intended balance of outcomes as program designers, and subsequently evaluators, are left to interpret relative policy priorities and decide what the optimal balance should be.

The main take-away from this exploration is that **there is not yet a clear process at the policy, program, or evaluation level to deal with trade-offs rooted in competing objectives in a meaningful and actionable way.**

**Complex Interrelationships of Policy Objectives and Programs**

As the energy system becomes more complex and programs seek to accomplish more policy objectives, there may be an increase in the use of energy programs as tools that need to work together toward the achievement of the broader policy objectives. This apparent trend changes the meaning of success from whether a program achieved its goals to whether it did its part in a larger system.

**State climate goals are one example of this dynamic.** Some states (and local governments) are setting climate-related targets and calling upon programs to help achieve them. When this charge remains directional (achieve GHG savings), program-specific evaluations can do little more than confirm that individual programs are designed to contribute to the overall objective and assess whether and how much a program may be contributing toward the climate goals. In theory, program-specific evaluations can also examine the average or marginal costs of those greenhouse gas reductions from the program activities, so they can be compared across programs. However, those evaluations do not answer important holistic questions, such as:

- Do the various programs charged with contributing to climate goals comprise a holistic and complete strategy? Can they be expected collectively to meet the established goals? Are they structured to complement each other? Is the combination of programs sound?
- Within any broader (stated or implied) cross-program strategy, what is the role each program is intended to play? Is it succeeding at doing its part toward the overall climate goal?

Answering these questions would require clarity on how the various programs are intended to work together to achieve the bigger purpose, information about the relative unit costs, and consistency across program-specific evaluations to ensure that collective progress and effectiveness toward the policy goal can be ascertained. (And when programs were not designed to work together, there is value in pointing to missed opportunities from disjointed programs and policies.) These are not questions evaluators alone can address, but they are important considerations for the policy goals to which evaluators can contribute.

**Energy affordability is another example of individual efforts that sometimes work together in a piecemeal fashion toward a broader goal.** Low-income programs exist to ensure that highly resource-constrained households can afford the energy they need for health, safety, and comfort. The energy industry has developed ways of measuring energy affordability using a well-defined metric of energy burden and set standards of how much energy burden is reasonable (generally accepted to be six percent of household income). This metric and standard positions evaluators to assess needs and evaluate low-income programs in meaningful ways.

However, as some states go further in seeking to assist resource constrained households, the lines between energy and other household needs get blurred. Policymakers appropriately consider household needs holistically. Households with resource constraints do so too. Nevertheless, programs to assist resource constrained households are siloed into energy programs, medical programs, food programs, housing programs, general assistance, and other initiatives. When policy makers are motivated to help households because of their general inability to make ends meet (beyond specifically energy-related needs), the programs available need to be viewed in their totality rather than individually.

**Changes in Values, Norms, and Public (and Personal) Sentiment**

One other change we are experiencing relates to values, norms, and public sentiment. Arguably, American society is being encouraged to think about previously apolitical topics in political terms.

**Gauging public sentiment.** The 2024 presidential election in the United States says something meaningful about individual preferences and priorities. There is a message here for energy programs that is worth hearing. It just is not completely clear what that message is. Last year, the United States elected a president who promised to “drill, baby, drill” and whose promises suggested that he would declare a national energy emergency to shift course away from climate-friendly policies and from renewable energy. A plurality of voters chose Donald Trump either because of those policies, in spite of them, or due to other priorities that suggest that climate and the recent direction of energy-related transitions were comparatively not important to them. Analysts seemed surprised at the level of support for Trump among voters from demographic groups that tend to support progressive policies related to energy, equity, and related issues.

Policy battles will play out in national and state arenas with federal initiatives likely reversing course to a degree from recent pathways that were sensitive to climate, encouraged decarbonization, and favored equity-related initiatives. Progressive-leaning states may dig into their positions or push their policy preferences even more aggressively. Utilities that are already transitioning toward lower carbon electric generation seem to plan to stay on course for now. State regulators debating the future of natural gas will continue their deliberations. Those dynamics seem clear, but there is an equally important question that is much harder to answer: What are voters telling us about their preferences and priorities and what does this mean for energy programs, design, and marketing when these same voters are asked to engage as consumers?

Key questions include:

- What aspects of recent and planned energy transitions have the support of the public? Which segments of the public?
- How will households and organizational decision-makers react when offered energy efficiency, renewable energy options, opportunities to electrify, or opt-in dynamic rates? Will the traditional participants in energy efficiency also engage in these other offerings, or will it be a different set of consumers?
- What benefits of those offers will appeal? What messages may trigger resistance, cognitive dissonance, or other negative reactions?
- What messages beyond dollars and cents will motivate or trigger resistance? How will value-based appeals related to climate or equity resonate? With whom?
- What will it take to “bring everyone along” or “not leave anyone behind” in the current political climate and in the context of politically acceptable messaging on all sides of the spectrum?

These are questions that individual programs will need to answer as they choose how to engage with market actors and attempt to spur voluntary action toward energy efficiency, demand response, climate-friendly choices, and so forth. However, it is inefficient and can be beyond the capacity of individual programs to figure out the answers, particularly when comprehensive answers require the perspectives of those who are not interested in the current offerings and are less likely to respond to standard research methods. Cross-jurisdiction, collaborative research may be the most cost-effective way to explore some of these topics and allow for shared resources to provide more comprehensive, deeper understanding.

**Personal values.** Energy evaluators are humans with personal value sets and political leanings. Informal discussions with energy evaluators can leave one with the impression that these professionals tend to care personally about energy efficiency, the climate, and recent social movements related to equity, and many of the companies active in the field express similar organizational values. While personal passion is a positive, it is important to be mindful of groupthink, to not ostracize unpopular perspectives, to understand how personal values can bias one's assessment of what is proper, and to be clear whose choice should guide the criteria by which programs are assessed.

## Considerations for a Mindful Evolution

There is much to reflect about as the practice of energy evaluation seeks to keep up with—and help guide—rapid transitions in the energy industry.

### What Is

In mindfulness, one seeks to be aware and accept "what is," and evaluators should be well suited to conduct such dispassionate observation about their own industry. It does require effort to focus one's attention on broader trends and needs in the midst of busy work schedules that tend to be focused on the specific details of narrowly defined project work.

This paper offers some thoughts about emerging trends and evolution in the energy industry that require thoughtfulness by energy evaluators. Discussion about these issues already happens at conferences and other gatherings that energy evaluators frequent, as well as in conference papers in which authors call out their own observations. There is a case to be made for expanding on this discussion in a more systematic format to more fully capture and discuss the needs energy evaluators are being called upon to address and opportunities to weigh in (even if we are not being asked yet).

### How to Respond

Evolving to meet the present and future needs for empirical insights is the more challenging task. This paper's author—informed by discussions with other evaluators—offers the following thoughts, while fully recognizing that this is one voice in what needs to be a larger discussion within the energy evaluation field. The broader field of energy evaluators may do well to focus on:

- Building on present strengths
- Borrowing from other evaluation practices
- Revisiting and challenging assumptions

**Build on present strengths.** The energy evaluation field already benefits from a good reputation, relies on empirical methods, and maintains well-established tools such as cost-effectiveness tests that have provided solid answers to the questions we have been asked since energy evaluation began. It is essential that the field maintain that reputation by staying true to empirical methods and treading cautiously when evaluation questions turn more subjective. That means putting aside our personal leanings on substantive matters and staying true to solid methodology.

That is not to say creativity and values-based thinking do not play a role in evaluation. They do when the aim of evaluation is to contribute directly to program design by offering solutions to problems. However, it is important to be clear for ourselves and with evaluation stakeholders when we are inserting ourselves into the process (and more likely to bring our own personal perspectives, experiences, and creative ideas) and when we are serving more of an accountability, verification, or independent assessment role (in which case conservative application of standards is substantially more important). There is a difference between program-informing research (including market research, formative evaluation, and embedded evaluation), and evaluation for the sake of accountability. They are all highly valuable, but they come with different "rules of engagement." As a result, it is important to be clear with stakeholders when we are engaged in which.

The energy evaluation field should also continue to use its existing tools and simultaneously build new ones. As noted, cost-effectiveness methodologies have been a strength of energy evaluation. The field could build on its existing tools to address new issues we are asked to address, such as measuring metrics related to carbon benefits and equity. A cooperative approach across the industry (and with involvement from stakeholders) would serve the field best, as it is ideal to have commonly accepted practices rather than black box proprietary ones or unique approaches in each jurisdiction. Perhaps a process similar to that used for the Uniform Methods Project would be a good model to follow.

**Borrow from other evaluation practices.** Government evaluation may hold some lessons for energy evaluators as the connection between programs and policy becomes more complicated. Government evaluators seem more comfortable speaking to policymakers as well as public sector agencies carrying out programs. In contrast, energy evaluators have a greater tendency to speak only to the programs and program administrators who hire them, taking the policy objectives as a given. However, this limits the ability of the policy process to self-correct when aspirational, vague, or conflicting goals result in ambiguity for programs and evaluators. Evaluation studies could help programs better play their role and contribute to the broader success of policy objectives if evaluators called out gaps, missed opportunities, or strategic misalignment in policy directives handed to programs (when they exist) and alternatives that would accomplish the goals better.

Relatedly, government evaluation tends to be rigorous in basing evaluation findings on clear criteria that are often taken from stated policy objectives. Conditions (i.e., what is) are compared to criteria (i.e., what should be), and those criteria are often linked to policy intent as well as best practices. Recommendations are based on this assessment. These tighter linkages shield evaluation findings from perceptions of bias and subjectivity, which is more important in government evaluation than it has been in energy evaluation. However, following this practice more closely would strengthen evaluation findings in the energy field and strengthen the rigor and credibility of comments on policy and cross-cutting issues.

**Revisit and challenge assumptions.** Finally, evaluators should always challenge assumptions, especially when addressing new topics and contexts. Given that political and social forces have impacted society recently just as energy industry initiatives are making novel asks of consumers, now may be a good time for a reality check on customer perspectives. The industry might consider broad-based research on the true priorities of individuals and households that are relevant to energy choices, markets, and programs, as well as test their appetite for the various new ways consumers are being asked to engage with energy. We could use a reality check on how consumers may respond to the substance of the energy-related asks, the way they are (or may be) presented, and the goals they are seeking to accomplish in the first place. Given the breadth of such research and the challenges the market research field is facing in obtaining responses from unengaged consumers who are key to the research, this is no trivial effort. It might be most efficiently done in a cross-cutting, collaborative fashion with results shared openly.

Further, various recent initiatives in the energy field rely on simplifying assumptions that may not be universally true. For example, electrification does not decarbonize until the electric generation is carbon free. Similarly, offering enhanced energy efficiency in geographies that are labeled as disadvantaged for non-energy reasons may not solve underlying issues or enhance equity to the degree that advocates might like to think. Evaluators ought to ask the tough questions and help to answer them, and that is true even if the evaluators are personally passionate about the topics being evaluated. Pointing out the weaknesses facilitates improvements and more effective solutions.

This is one set of observations based on one evaluator's observations and informed by informal conversations with peers. Hopefully, they offer some food for thought. Zooming out from these perspectives, it seems fair to say the energy field is evolving substantially, and the practice of energy evaluation needs to evolve with it in various ways while staying true to its core principles.

## Next Steps

This paper is intended to be the start of a conversation rather than "the roadmap" for how energy evaluation keeps up with a transitioning energy industry and changing informational needs. While individual evaluation providers likely are monitoring industry developments in an effort to position themselves in a competitive landscape, there is simultaneously a need for open, transparent, and collective discussions across the industry to help the energy evaluation field respond to the needs of the moment and the future. Here are some suggestions for the field:

- Industry organizations (like the International Energy Program Evaluation Conference) could create structured opportunities for energy evaluation practitioners to develop a common understanding of how the energy evaluation field can best contribute to sound and effective programs and policies amidst the evolution in the energy industry. These efforts could be open meetings at industry events, facilitated virtual "town hall" conversations, or the development of a working group of volunteers.
    - ⇒ The discussions could be organized similarly to this paper, beginning with discovery of what is changing and what challenges and opportunities the energy evaluation field needs to address mindfully.
    - ⇒ Some of the discussions need to include thought-leading stakeholders of energy evaluation services.
    - ⇒ There may also be a benefit in collective talking points for the energy evaluation field to be communicated to stakeholders whose engagement with energy evaluation is more transactional. Communicating our role better to those who develop scopes of work and manage evaluation projects can help create a common understanding, unlock strategic opportunities to better inform energy policy and initiatives, and prevent misplaced expectations.
- Discussions among energy evaluators should include a focus on identifying approaches and tools we are missing for which industry-accepted standards and practices would be helpful (to ensure consistency and minimize inefficient re-inventing of the wheel).
    - ⇒ Once needs for new approaches and tools are identified, one or more entities could lead the development of industry-accepted frameworks. In the past, the U.S. Department of Energy has played such a role. If DOE is not available or appropriate, other possibilities include states that are active users of evaluation and research or self-funding within the energy evaluation field.
- If these steps are put in motion, IEPEC could consider including tailored working sessions and updates in its 2027 conference agenda.

## References

AEA (American Evaluation Association). 2014. "What is evaluation?" Retrieved from https://www.eval.org/Portals/0/What is evaluation Document.pdf.

Bensch, Ingo. 2024. "Balancing Equity and Energy Objectives: Dealing with the Trade-Offs." In Proceedings of the 2024 ACEEE Summer Study on Energy Efficiency in Buildings. Washington, DC: ACEEE.

CPUC (California Public Utility Commission). 2001. California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects. San Francisco: CPUC. www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/Utilities_and_Industries/Energy_-_Electricity_and_Natural_Gas/CPUC_STANDARD_PRACTICE_MANUAL.pdf.

Efficiency Value Organization. 2022. International Performance Measurement and Verification Protocol. Washington, DC: Efficiency Value Organization. https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp

Energy Equity Project, 2022. Energy Equity Framework: Combining data and qualitative approaches to ensure equity in the energy transition. University of Michigan – School for Environment and Sustainability.https://energyequityproject.org/wp-content/uploads/2022/08/220174_EEP_Report_8302022.pdf

GAO (Government Accountability Office). 2021. Program Evaluation: Key Terms and Concepts. Washington, DC: United States Government Accountability Office. https://www.gao.gov/assets/gao-21-404sp.pdf.

GAO (Government Accountability Office). 2024. Government Auditing Standards: 2024 Revision. Washington, DC: United States Government Accountability Office. https://www.gao.gov/assets/d24106786.pdf

Li, M.; Haeri, H.; Reynolds, A. 2018. The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures. Golden, CO; National Renewable Energy Laboratory. NREL/SR-7A40-70472. http://www.nrel.gov/docs/fy18osti/70472.pdf

OMB (Office of Management and Budget). 2021. Memorandum for Heads of Executive Departments and Agencies (M-21-27). Washington, DC: Executive Office of the President. https://bidenwhitehouse.archives.gov/wp-content/uploads/2021/06/M-21-27.pdf