DOE/EE-2087

Performance Assessments of Demand Flexibility from Grid-Interactive Efficient Buildings: Issues and Considerations

July 2020

The State and Local Energy Efficiency Action Network is a state and local effort facilitated by the federal government that helps states, utilities, and other local stakeholders take energy efficiency to scale and achieve all cost-effective energy efficiency by 2020.

Learn more at www.seeaction.energy.gov

If this document is referenced, it should be cited as:

State and Local Energy Efficiency Action Network. (2020). *Performance Assessments of Demand Flexibility from Grid-Interactive Efficient Buildings: Issues and Considerations*. Prepared by: Steven R. Schiller, Lisa Schwartz, and Sean Murphy, Lawrence Berkeley National Laboratory.

FOR MORE INFORMATION

Regarding *Performance Assessments of Demand Flexibility from Grid-Interactive Efficient Buildings: Issues and Considerations,* please contact:

Lisa Schwartz
Lawrence Berkeley National Laboratory
LCSchwartz@lbl.gov

Regarding the State and Local Energy Efficiency Action Network, please contact:

Johanna Zetterberg
U.S. Department of Energy
johanna.zetterberg@hq.doe.gov

Acknowledgments

The work described in this study was funded by the U.S. Department of Energy's (DOE's) Building Technologies Office, Office of Energy Efficiency and Renewable Energy, under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231. This study was prepared under DOE's authority to encourage and facilitate the exchange of information among State and local governments with respect to energy conservation and energy efficiency, and provide technical assistance on such matters. This study was specifically prepared for the use and dissemination of the State and Local Energy Efficiency (SEE) Action Network, a DOE program that DOE and the U.S. Environmental Protection Agency (EPA) facilitate, to offer resources and technical assistance to state and local decision makers as they provide low-cost, reliable energy to their communities through energy efficiency. The purpose of SEE Action Network is not to provide advice or recommendations to DOE or the Federal government.

The authors thank our DOE program managers, Monica Neukomm and Johanna Zetterberg, and the following individuals for reviewing all or part of the draft report: David Nemtzow, John Agan, Ryan Fedie, Dale Hoffmeyer, and Sarah Zaleski (DOE); Joe Bryson (EPA); Debra Lew (Debra Lew, LLC); Jingjing Liu, Bruce Nordman, and Mary Ann Piette (Berkeley Lab); David Narang (National Renewable Energy Laboratory); Rodney Sobin (National Association of State Energy Officials); Elizabeth Titus, Giselle Procaccianti, and Sue Coakley (Northeast Energy Efficiency Partnerships); Henry Yoshimura (ISO New England); Mia Adams (Midcontinent ISO); Carmen Best (Recurve); Brian Kooiman (OhmConnect); Brenda Chew (Smart Electric Power Alliance); Tim Woolf (Synapse Energy Economics); Julie Michals (E4theFuture); and Tom Eckman (Berkeley Lab consultant). We also thank other experts we interviewed (Appendix A) and SEE Action Network Executive Committee members who provided feedback on draft materials for the report.

All opinions, errors, and omissions remain the responsibility of the authors. All reference URLs were accurate as of the date of publication.

Other Reports in This Series

- ***Introduction for State and Local Governments***: Describes grid-interactive efficient buildings in the context of state and local government interests; highlights trends, challenges, and opportunities for demand flexibility; provides an overview of valuation and performance assessments for demand flexibility; and outlines actions that state and local governments can take, in concert with utilities, regional grid operators, and building owners, to advance demand flexibility
- ***Determining Utility System Value of Demand Flexibility from Grid-Interactive Efficient Buildings***: Describes how current methods and practices that establish value to the electric utility system of investments in energy efficiency and other distributed energy resources (DERs) can be enhanced to determine the value of grid services provided by demand flexibility resources

In addition, DOE offers a fact sheet, overview, and series of technical reports with more information on technologies and strategies for grid-interactive efficient buildings: https://www.energy.gov/eere/buildings/grid-interactive-efficient-buildings.

Abstract

This SEE Action Network report explains basic concepts and fundamental considerations for assessing the actual demand flexibility performance of buildings participating in demand flexibility programs and responding to time-varying retail rates. Demand flexibility is the capability of distributed energy resources (DERs) to adjust a building's load profile across different timescales. Assessments determine the timing, location, quantity, and quality of grid services provided.

The results can be used for financial settlements and to improve performance of demand flexibility, support its consideration in resource potential studies and electricity system planning, and contribute to cost-effectiveness evaluations.

While practitioners and regulators regularly find opportunities to improve performance assessments of demand-related services, to a large degree current best practices are sufficient for basic service offerings as demand flexibility is implemented today. However, advances in assessment practices will be required in a future with grid-interactive efficient buildings (GEBs) that provide continuous demand flexibility by integrating multiple DERs and flexibility modes (load shed, load shift, modulate, and generate). Example practices include using new baseline constructs—including whether baselines are even required for certain building flexibility modes or configurations, deploying more advanced metering and analytics, further developing cybersecurity standards, improving communication standards for increased interoperability, and establishing performance metrics and assessment procedures for load modulation as it is more fully defined and implemented.

This report provides information on prioritizing and designing performance assessment elements for demand flexibility programs and time-varying retail rates, assessment protocols, and research and development needs. Additional research sponsored by DOE's Building Technologies Office[1] offers more detailed technical information on assessing performance of demand flexibility, including definition of performance metrics.

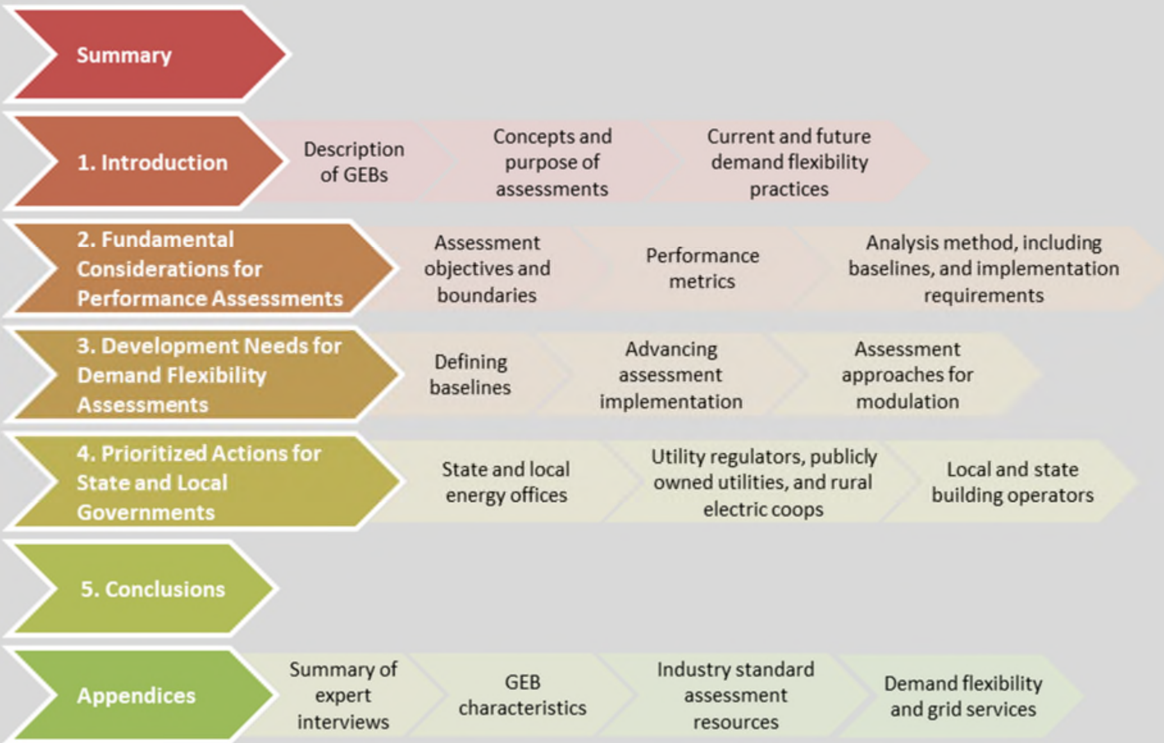
Target Audience

The target audience for this report is state and local governments, including utility regulators. Utilities, regional grid operators, energy service providers including DER aggregators, building system designers, and building owners, operators, and occupants also can use the information to improve demand flexibility. In addition, this report may aid researchers in analyzing demand flexibility frameworks and metrics in support of developing and testing assessment protocols.

[1] See https://www.energy.gov/eere/buildings/grid-interactive-efficient-buildings.

How to Use This Report

This report synthesizes fundamental considerations for assessments of demand flexibility provided by individual buildings today and in the future. The report is organized as follows.



- The *Summary* section provides major findings and recommendations.
- *Section 1* describes GEBs and their characteristics, defines the concepts and purposes of demand flexibility assessments, summarizes current demand flexibility practices, and describes expected advancements both in capabilities of buildings to provide demand flexibility and in assessment practices.
- *Section 2* presents five fundamental considerations associated with performance assessments of demand flexibility: (1) assessment objectives, (2) assessment boundaries, (3) performance metrics (both number and quality of metrics), (4) analysis methods (including baselines), and (5) assessment implementation requirements (for example, for data collection and privacy).
- *Section 3* builds on these five fundamental considerations to describe development needs for assessments in a future with buildings providing more complex grid services. The three areas of focus are: (1) developing new baseline constructs; (2) implementing assessments, with attention to additional metering needs, interoperability and communication standards, privacy, cybersecurity, and independent evaluators; and (3) assessing modulation as a demand flexibility service.
- *Section 4* summarizes ways state and local governments and others can support implementing, standardizing, and enhancing assessments to support cost-effective services from demand flexibility.
- *Appendix A* summarizes our interviews with experts. Other appendices provide references and additional information on select topics—GEB characteristics, industry-standard measurement and verification approaches, and demand-side management strategies and grid services.

Acronyms

AGC	automatic generation control
AMI	advanced metering infrastructure
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BAS	building automation system
CPP	critical peak pricing
CVR	conservation voltage reduction
DER	distributed energy resource
DOE	U.S. Department of Energy
EMIS	energy management information system
EM&V	evaluation, measurement, and verification
EV	electric vehicle
FERC	Federal Energy Regulatory Commission
GEB	grid-interactive efficient building
GMLC	Grid Modernization Laboratory Consortium
HVAC	heating, ventilating, and air-conditioning
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet-of-Things
ISO	independent system operator
M&V	measurement and verification
NIST	National Institute of Standards
PUC	public utility commission
PV	photovoltaic
RCT	randomized control trial
RD&D	research, development, and demonstration
RTO	regional transmission organization
TOU	time of use
T&D	transmission and distribution

Glossary[2]

Demand flexibility: Capability of DERs to adjust a building's load profile across different timescales; energy flexibility and load flexibility are often used interchangeably with demand flexibility.

Demand response: Change in rate of electricity consumption in response to price signals or specific requests of a grid operator.

Distributed energy resource (DER): A resource sited close to customers that can provide all or some of their immediate power needs and/or can be used by the utility system to either reduce demand or provide supply to satisfy the energy, capacity, or ancillary service needs of the grid.

Grid-interactive efficient building (GEB): An energy-efficient building that uses smart technologies and on-site DERs to provide demand flexibility while co-optimizing for energy cost, grid services, and occupant needs and preferences, in a continuous and integrated way.

Grid services: Services that support the generation, transmission, and distribution of electricity. This report focuses on grid services that can be provided by grid-interactive efficient buildings.

Impact evaluation: A performance assessment of multiple buildings in a program or tariff to determine its impacts, such as energy or demand savings.

Interoperability: The capability of two or more networks, systems, devices, applications, or components to externally exchange and readily use information securely and effectively.

Metrics: Numbers, or other forms of information describing the process of interest, that indicate how the process is performing. Metrics provide a basis for suggesting or making improvements to the process.

Smart technologies for energy management: Advanced controls, sensors, models, and analytics used to manage DERs. GEBs are characterized by their use of these technologies.

[2] Definitions are from Neukomm et al. 2019, except for impact evaluation, interoperability, and metrics.

Table of Contents

List of Tables

List of Figures

Summary

Demand flexibility in buildings supports electricity system reliability and resilience, energy affordability, integration of new generating resources and loads, and other state and local energy goals. By contributing grid services needed for these purposes, demand flexibility can help advance a jurisdiction's energy-related policies such as integrating energy efficiency with other distributed energy resources (DERs), hardening critical energy infrastructure, reducing peak demand, mitigating climate change, achieving renewable energy targets, and electrifying transportation and targeted building loads.

Performance assessments of demand flexibility determine the timing, location, quantity, and quality of grid services provided. Such assessments are common for financial settlements and have other important applications (Figure S-1). For utilities, regional grid operators, and utility regulators, assessments provide confirmation that buildings can reliably and consistently provide demand flexibility, critical to its broad acceptance as a grid resource. For building owners, operators, and occupants, assessments help optimize building performance, provide confidence in the benefits they will receive from demand flexibility (e.g., lower energy costs, payments for performance), and demonstrate acceptable non-energy impacts (e.g., building maintains comfort standards). Assessments also can reveal positive non-energy impacts, such as improved equipment performance through better building monitoring and controls and higher building resale value through lower net energy costs. In addition, comprehensive assessments provide state and local governments data they need to advance demand flexibility in support of their broader energy goals. In these ways, all stakeholders benefit from information that assessments provide for planning, designing, and implementing demand flexibility cost-efficiently.

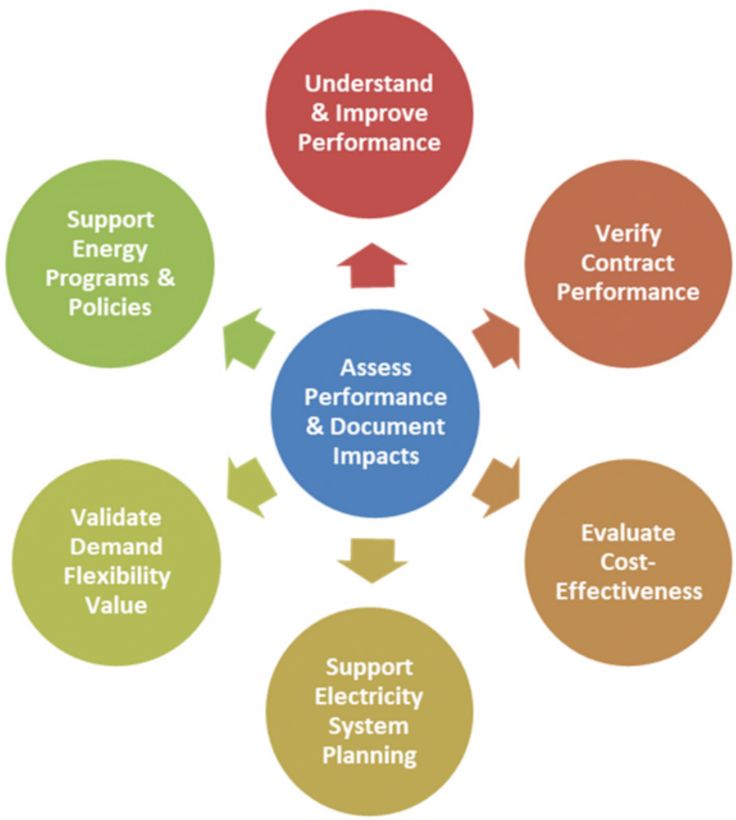


Figure S-1. Multiple Values for Performance Assessments

Assessments of demand flexibility serve several purposes for various audiences.

DEMAND FLEXIBILITY

Demand flexibility is the capability of DERs to adjust a building's load profile across different timescales. Demand flexibility (or load flexibility), integrated with energy efficiency, is the core characteristic of grid-interactive efficient buildings. The potential impacts are significant. Buildings account for 75% of U.S. electricity consumption and a comparable share of peak power demand.

Source: U.S. Energy Information Administration 2019

This report synthesizes basic concepts and fundamental considerations and identifies development needs associated with assessing performance of *individual* buildings (residential, commercial, institutional) participating in dispatchable demand flexibility programs. Individual buildings are the building blocks for assessing *multiple* buildings aggregated for programs. Dispatchable demand flexibility programs provide financial incentives to consumers that change electricity demand in response to events called by a utility, regional grid operator, or DER aggregator for reliability or economic reasons.

The report includes some discussion of assessment methods for multiple buildings, known as *impact evaluations,* in the context of assessing time-varying retail rates that also encourage demand flexibility. In addition, performance assessments may be conducted for voluntary demand flexibility efforts encouraged by government entities or utilities during times of electricity system stress. While not the subject of this report, market evaluation—another type of assessment—can be a valuable tool for increasing adoption of demand flexibility, to assess progress toward market transformation and other market-related objectives.

In addition, the report identifies ways that state and local governments and other stakeholders can support meeting assessment needs in order to advance demand flexibility for reliable grid services. Government agencies, utilities, regional grid operators, DER aggregators, energy service providers, building system designers, and building owners, operators, and occupants can use the information in this report to improve their understanding of the role of assessments and as a starting point for defining assessment procedures and requirements, such as those associated with data collection and analysis.

Grid-interactive efficient buildings (GEBs) use smart technologies, including advanced controls[3] and sensors, to actively manage DERs—energy efficiency, demand response, distributed generation and storage, and managed electric vehicle (EV) charging—to optimize energy use for grid services, occupant needs and preferences, cost reductions, and other purposes in a continuous and integrated way. In addition to providing ongoing reductions in energy use through energy efficiency, a building's load profile can be adjusted across different timescales using four demand flexibility modes, individually or in combination (see Section 1.1 for definitions):

- Load shed
- Load shift
- Modulate
- Generate.

In the future, buildings will use multiple DERs and demand flexibility modes to respond to grid needs quickly, even within seconds or sub-seconds, potentially providing continuous demand flexibility. These changes will enable buildings to provide additional grid services, but will require advances in demand flexibility performance assessments.

[3] See Fernandez et al. 2017.

Effective performance assessments use data to quantify the amount and quality of demand flexibility provided by a building with respect to predefined performance metrics. For example, assessments can indicate the average amount of load shed or shifted, in kilowatts (kW), over a given time period and how quickly a building's shedding or shifting of load ramped up to provide a sustained change in demand. Beyond short-term uses of such information, such as for financial settlements, assessments support resource potential studies and electricity system planning, including evaluating cost-effectiveness of demand flexibility compared to other options for meeting generation and transmission and distribution (T&D) needs.[4] Demand flexibility assessments also can adopt integrated or holistic approaches that take into account a jurisdiction's related energy programs and policy goals.

Five fundamental considerations define a demand flexibility (and energy efficiency) assessment:

1. *Assessment objectives*–What information is the assessment intended to provide and how will the information be used?

2. *Assessment boundary*–At what level will performance be assessed—whole (individual) building or building system or equipment level, by DER, and/or by demand flexibility mode (e.g., load shed, load shift, modulate, or generate)? Or will the assessment boundary encompass multiple buildings, perhaps defined by their location (e.g., all buildings served by a single substation on the electric grid)?

3. *Performance metrics*–What metrics will be assessed and how will be they defined? What data will be required and at what temporal granularity (e.g., sub-seconds to hours for data measurement frequency) to calculate metric values?

4. *Analysis methods*–How will metrics be calculated and with what expectations for certainty? Will baselines be used and, if so, how will they be defined?

5. *Assessment implementation requirements*–What are the requirements with respect to data collection, privacy, cybersecurity, and reporting—particularly timing of reporting (e.g., in real time, within hours or days)? What entities will conduct the assessments? What is the duration of assessments (performance period covered)?

The most critical step in designing an assessment is determining appropriate performance metrics. Metrics are numbers, or other forms of information (e.g., categorical values), describing a process in a manner that indicates how well it is performing. Metrics provide a basis for making process improvements. For assessing demand flexibility performance in individual buildings, metrics may encompass four dimensions:

- Quantity and timing of demand flexibility provided (e.g., amount of demand reduction during defined period in kW or kilowatt-hours (kWh), the most common metric for demand flexibility today)

- Quality of demand flexibility provided (e.g., speed of achieving desired demand change or persistence of desired demand flexibility over long periods of time)

- Attribution of impacts to equipment, DER, flexibility mode, and/or building location (performance assessment boundary)[5]

    - Individual equipment (e.g., chillers) and systems (e.g., lighting)

    - Individual DERs

[4] See National Efficiency Screening Project, *National Standard Practice Manual for Benefit-Cost Analysis of Distributed Energy Resources*, forthcoming.

[5] For performance assessments of multiple buildings, the assessment boundary is expanded beyond the individual building, equipment, or system to encompass all buildings—or all of the specific equipment or systems of interest (e.g., all air conditioners in buildings on a specific time-varying retail tariff)—as well as location of buildings on the electricity grid.

- Individual flexibility modes
- Location of impacts on the electricity grid

- Impacts on owners and occupants, including energy cost savings and non-energy impacts such as comfort, health, and productivity.

For some demand flexibility modes and grid services provided, performance metrics may simply be based on direct measurements, such as the amount of electricity provided to the grid from a building with on-site generation. Alternatively, performance metrics may be based on whether the building's demand for electricity is within desired parameters, such as a desired load shape (e.g., lower demand during certain hours and higher demand during other hours).

But for most types of demand flexibility and grid services provided, direct measurements must be compared to other quantities—typically a counterfactual scenario (commonly referred to as the *baseline*)—to understand the quantity of grid services provided. For example, the amount of load shed during a specific time period is equal to the difference between the actual load of the building and a counterfactual scenario, defined as the load that would have occurred in the absence of the subject utility demand flexibility program or time-varying retail rate. Defining and determining counterfactual scenarios for most metrics is a major component of assessment analytics, particularly for dispatchable demand flexibility programs which typically rely on historical data for building electricity demand to define baselines. In contrast, the counterfactual scenario for time-varying retail rates, such as time-of-use (TOU) pricing, can be defined by a control group[6] without attribution for individual building performance, because no settlement process is involved for individual participants. Figure S-2 presents a hierarchy of analysis methods for demand flexibility assessments.

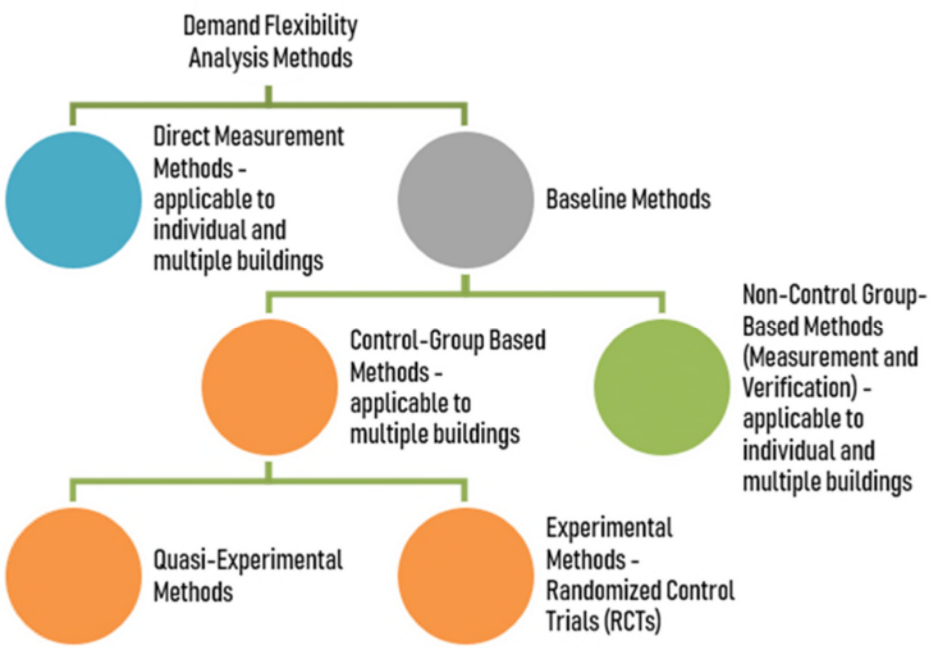


Figure S-2. Hierarchy of Analysis Methods for Demand Flexibility Assessments

Analysis methods may use direct measurement methods or baseline methods. Because matching is feasible only in the context of multiple buildings, control group methods do not apply to individual buildings.

[6] Control groups consist of buildings that are not on the subject tariff, or do not participate in the subject program, but whose characteristics are very similar to those buildings that participate in the tariff or program.

Impacts such as demand (kW) change and energy (kWh) savings, and thus metrics, are quantified in the same manner regardless of the state's electricity market structure—vertically integrated utilities or centrally organized wholesale electricity markets (or a combination). Conceptually, all performance metrics may be assessed at the whole building or individual system or equipment level, or for all buildings participating in a demand flexibility program or time-varying retail rate. Performance metrics may differentiate between impacts associated with energy efficiency or the four modes of demand flexibility (load shed, load shift, modulate, and generate) and which DERs are providing such flexibility. Still, metrics of greatest interest are associated with demand flexibility performance of a whole building or an aggregation of buildings.

Demand flexibility assessments can build on existing approaches for performance verification, such as demand response measurement and verification (M&V) protocols for utility programs and wholesale electricity markets. The fundamentals of assessing demand reductions from energy efficiency, load shed, and generation have been in place for decades, are well-established, and can serve as the basis for load shifting and modulation assessments. Particularly applicable are advanced M&V practices that have adopted the use of smart meter data and automated analytics. For example, demand flexibility assessments can adapt existing practices related to metering and data quality standards, measurement protocols, counterfactual scenario definitions, and use of independent third parties. Further, at least as a starting point, assessments can take advantage of approaches to data privacy and cybersecurity, as well as installation of building automation systems (BAS) and advanced metering infrastructure (AMI). Where deployed, BAS and AMI support data collection and sharing and may significantly ease implementation and assessment of demand flexibility.

Advanced M&V (also called M&V 2.0) can help realize the promise of buildings as grid assets in two important ways. First, it leverages standardized data formats and the finer timescale of smart meter data. Second, automated analytics enable processing of larger volumes of data at high speeds to support fast-response modulation implementation and assessments. Combined with conventional impact evaluations, used regularly to assess load shedding associated with time-varying rates, a strong base of existing practices can be drawn upon to assess demand flexibility as currently implemented.

In order to meet the full potential of GEBs to integrate multiple DERs and flexibility modes, however, existing assessment approaches will need to be modified or new approaches will need to be developed to assess demand flexibility performance—for example, in the context of:

- Buildings providing continuous or near-continuous demand flexibility
- Engagement of multiple demand flexibility modes in an integrated manner
- Load modulation in sub-seconds to seconds, autonomously providing grid stability and balancing services
- Increased use of combinations of DERs, such as on-site generation to charge batteries in concert with load shed and shift
- Integrated whole-building system approaches to providing grid services, as well as demand flexibility at the individual end-use level or individual device level
- Managed EV charging
- Reducing complexity associated with multiple DERs, demand flexibility modes, and program and rate designs by providing simplified approaches for consumers and other market participants.

These examples imply more sophisticated assessments of demand flexibility.

Table S-1 summarizes five drivers of development needs for future demand flexibility assessments and three priority development needs with respect to baseline constructs, implementation practices and infrastructure, and modulation.

Table S-1. Drivers of Priority Development Needs for Demand Flexibility Assessments

Drivers	Priority Development Needs
Increased use of shift and shed flexibility modes in combination creates overlapping baselines. Buildings providing multiple grid services, potentially in response to multiple programs offered by the servicing utility and/or regional grid operator, can create challenges for defining baselines. Increased number of events under dispatchable demand flexibility programs, including the potential for continuous demand flexibility, creates additional baseline challenges.	New baseline constructs (or replacements for baselines) will be needed in a future where one or more of these conditions occur.
Expanded objectives to include understanding contributions to overall demand flexibility performance of individual equipment or a particular building system (e.g., water heating system), DER, or flexibility mode.	Improvements in assessment implementation practices (e.g., communication and cybersecurity protocols) and infrastructure (e.g., metering and advanced M&V) will be needed to address, among other issues, the large volumes of short interval data that result when many DERs and GEBs are providing demand flexibility at regular intervals and the need for analyzing impact information quickly to support grid operation.
Demand-side modulation (ability to balance power supply/demand or reactive power draw or supply autonomously, within seconds to sub-seconds) is not a common practice today, although batteries are beginning to provide this service. Assessment practices for demand-side modulation will need to be established as this demand flexibility strategy is further developed with respect to services provided (e.g., frequency and voltage support), whether services are provided continuously or just in response to events, and whether implementation is in response to signals from the utility or regional grid operator or in response to direct measurement of grid conditions (e.g., voltage and frequency at the building/grid interface).	Assuming the modulation flexibility mode becomes more common—particularly with respect to being continuous, autonomous, and providing very fast response—metrics, measurement requirements, analysis protocols, and implementation protocols for modulation will need to be established.

New assessment approaches may best take the form of standardized protocols that can reduce costs and increase consistency and credibility of assessment findings. Such approaches likely will be based on frameworks that emphasize reasonable costs for assessing performance by considering meter-based analyses as an integral element of program implementation and resource planning. Inevitably, methods will emphasize consistent, simplified tracking and reporting that enables continuous and rapid feedback and opportunities for improvement in demand flexibility performance.

Research, development, and demonstration (RD&D) is required to address these needs. State and local governments can support such RD&D in several ways:

- Lead by example by conducting assessments of demand flexibility in their own public buildings.
- Encourage or require performance assessments for buildings participating in demand flexibility programs they operate or regulate, and for buildings on time-varying retail rates.
- Catalog and consider adopting best assessment practices, consistent with the jurisdiction’s policies and regulations.

- Share and support advances in assessment practices related to performance metrics, analysis methods, baselines, and implementation approaches for determining demand flexibility impacts.
- Disseminate assessment results and the metrics, data, analysis methods, and implementation strategies used. Leveraging and sharing data results in increased understanding and continuous improvement of demand flexibility performance, informs DER potential studies and electricity system planning based on verified performance, and secures confidence in performance assessment practices and thus demand flexibility as a grid resource, all in the most cost-effective manner.

Public utility commissions (PUCs) and state energy offices also can play an important role by providing guidance on performance metrics and standardized protocols for verifying buildings' demand flexibility in response to grid signals in real time, as well as over longer time periods. Standardization can reduce costs and increase consistency and credibility of assessment findings. States can take these actions in collaboration with utilities and regional grid operators, which write the assessment rules for verifying demand flexibility performance of program participants.

Other opportunities, including those for PUCs and state energy offices, include updating assessment strategies, such as baselines and qualifications criteria for entities that conduct and report assessment results, and encouraging development and adoption of standard, cost-efficient protocols for communication, data privacy, and cybersecurity. Additional actions include improving access to data necessary to conduct assessments. That may include facilitating investments in metering infrastructure, such as AMI and BAS with real-time measurement capability and built-in, two-way communication capability consistent with established standards and protocols for interoperability, cybersecurity, and data privacy.

1. Introduction

This report explains basic concepts, fundamental considerations, and development needs for assessing demand flexibility performance of buildings participating in demand flexibility programs and time-varying retail rates. Such assessments determine the quantity, timing, location on the grid, and quality of grid services provided.

Our research for the report employed three reinforcing approaches. First, we conducted a literature review on current assessment practices for each demand flexibility mode and technical and operational considerations for GEBs. Second, we collaborated with other Berkeley Lab researchers who are developing demand flexibility metrics. Third, we interviewed more than a dozen other experts, including representatives of utilities, regional grid operators, DER aggregators, and evaluation companies.[7] We then compiled and synthesized the results of our research and interviews to prepare this report and iterated with other lab and external experts on our findings and presentation.

1.1. Grid-Interactive Efficient Buildings and Demand Flexibility

Buildings have served as energy assets for decades, providing load management services and generating electricity on-site, with owners and tenants often participating in programs that state and local governments run or oversee. Recent innovations in building technologies and internet and grid connectivity open up new prospects for buildings to provide a broader suite of grid services. In the future, many residential and commercial buildings will continuously manage loads and DERs—energy efficiency, demand response, distributed generation and storage, and EV charging—to better serve the needs of building owners and occupants, electric utility systems, and regional grids. Demand flexibility from GEBs optimizes energy use for grid services, occupant needs and preferences, and cost reductions in a continuous and integrated way (Figure 1-1).

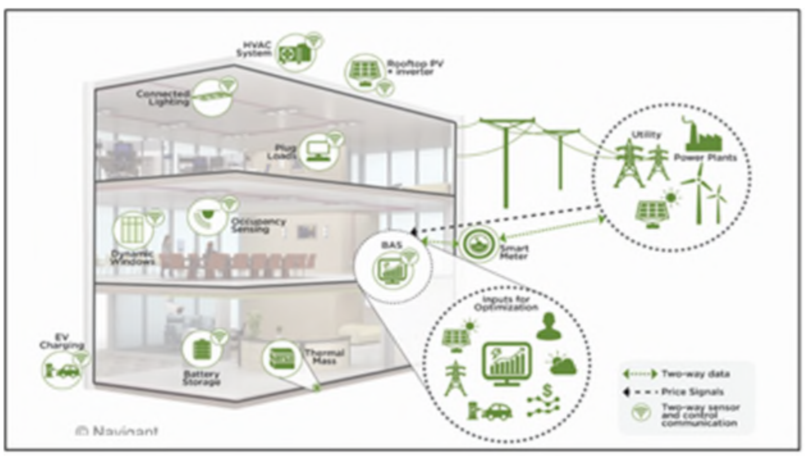


Figure 1-1. Example Commercial Grid-Interactive Efficient Building[8]

Advanced building technologies—including heating, ventilating, and air-conditioning (HVAC) controls, connected lighting, dynamic windows, occupancy sensing, thermal mass, and on-site generation, such as distributed photovoltaic (PV) and combined heat and power systems—are optimized to meet occupant and grid needs. A BAS responds to inputs such as sensors, weather forecasts, and price or event signals from utilities, regional grid operators, and third-party service providers.

[7] Appendix A summarizes the results of those interviews.

[8] Neukomm et al. 2019.

- **Efficiency:** Ongoing reduction in energy use while providing the same or improved level of building function.[9] Energy efficiency resources differ from demand flexibility modes in that they do not actively adjust loads and in general represent a set level of load reduction.[10]
- **Load shed:** Ability to reduce electricity use for a short time period.
- **Load shift:** Ability to change the timing of electricity use. In some situations, a shift may lead to changing the amount of electricity that is consumed.[11]
- **Modulate:** Ability to balance power supply/demand or reactive power draw/supply autonomously[12] (within seconds to sub-seconds) in response to a signal from the grid operator during the dispatch period.
- **Generate:** Ability to generate electricity for on-site consumption and dispatch electricity to the grid in response to a grid signal. Batteries can be considered in tandem, as they allow variable sources of power (e.g., solar PV) to be dispatched to some extent coincident with price levels or system peak conditions.

GEB CHARACTERISTICS SUPPORT ASSESSMENTS

GEBs use smart technologies for energy management and communication with the grid to provide demand flexibility for the electricity system. These two characteristics, smart and connected, also directly support cost-effective demand flexibility assessments.

Being *smart* implies buildings have metering and equipment status information as well as built-in analytical software that can be used to calculate (quantify) metrics of interest, all of which reduces the incremental cost of conducting assessments. Being *connected* supports real-time communication of performance information to grid operators, utilities, aggregators, and building operators. Appendix B describes how GEB characteristics support assessments.

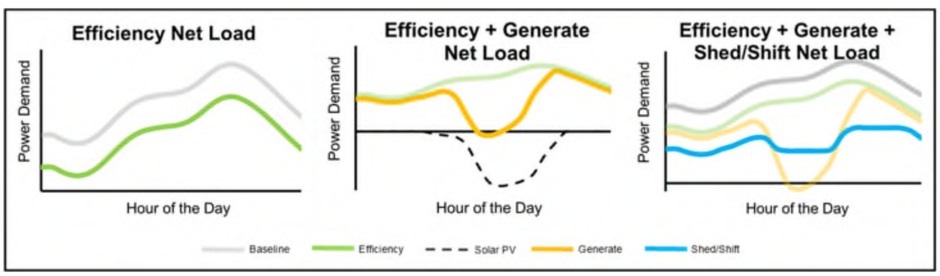


Figure 1-2. Daily Average Load Profiles for a Grid-Interactive Efficient Building[13]

Left: Energy efficiency alone pushes down the load curve. *Middle:* Energy efficiency plus distributed generation (in this case, solar PV) reduce overall energy use, but the building's peak load coincides with utility system peaks. *Right:* Adding load shedding and shifting flattens the building load profile, providing the greatest support to the grid.

[9] This would have the greatest impact for the grid during high-cost periods and minimize utilization of costly generation resources. Some efficiency measures, such as lighting controls, enable active management of efficiency resources and additional grid services.
[10] Federal Energy Regulatory Commission (FERC) 2014.
[11] Load shift in this report focuses on intentional, planned shifting such as minimizing demand during peak periods, taking advantage of low electricity prices, or reducing renewable energy curtailment. For some technologies, load shed can lead to some load shifting at times.
[12] This report includes two modulation response options (see Section 3.3): Autonomous—when devices, such as inverters, within a building directly monitor grid conditions (e.g., frequency) and respond by adjusting power output from distributed generation/storage systems or by reducing building electricity consumption. Automatic—when a building's BAS or individual devices receive a control signal from the grid operator and respond by adjusting load or power output from distributed generation/storage systems or reducing electricity consumption.
[13] Neukomm et al. 2019.

1.2. Assessing Demand Flexibility: Evaluation Use Case

This report primarily addresses assessment of demand flexibility performance (impacts) for an individual building[14] after adoption of demand flexibility measures or time-varying retail rates. Such assessments are one type of evaluation that may be conducted for demand flexibility. They use performance data to quantify and document metrics that indicate actual demand flexibility performance.

Metrics are numbers, or other forms of information (e.g., categorical values), that indicate the performance of the process of interest. Table 1-1 describes four use cases for demand flexibility metrics. Metrics for each use case are likely to be similar, if not identical. A distinguishing factor between use cases is that metrics are assessed at different points in time (before or after occurrence of grid events and implementation of demand flexibility modes), cover different time periods (a few minutes ahead of the demand flexibility activity to years into the future), and with different degrees of accuracy (projected versus actual performance). Thus, what distinguishes the evaluation use case from the other use cases is assessment of *actual* performance versus *projected* performance.

Table 1-1. Use Cases for Demand Flexibility Metrics[15]

Use Cases	Target Audience	Time Horizon	Aggregation Level	Metrics Examples	Example Applications
Planning	Utilities, regional grid operators	1–3 years ahead	Individual large building or groups of buildings	Rough estimates of load change during program periods	Program recruitment or tariff design
Operations	Aggregators, regional grid operators	Day ahead to 5 minutes ahead	Individual large building or groups of buildings	Estimate of desired or expected change in quantity of load during program implementation	Demand flexibility scheduling
Evaluation (including Performance Assessments)	Aggregators, regional grid operators, utilities, building owners or managers	Historic to near real time	Individual building or groups of buildings (aggregation)	Actual load change during program periods	Assessments of actual performance, support for improved planning, operations and design
Design	Building owners, designers, state regulators (for program design)	2–5 years ahead	Individual buildings or complexes	Energy cost savings for building owner or occupants; grid reliability and emissions	Building design/retrofit

[14] While the focus of this report is on buildings, the concepts and approaches are broadly applicable to other energy-consuming applications such as industrial facilities and street lighting.

[15] Adapted from Liu 2019.

1.3. Importance of Demand Flexibility Assessments

Planning processes estimate the value to the electric utility system of potential demand flexibility in buildings. Performance assessments verify whether expected demand flexibility is actually provided and are critical to confirming its value to the electricity system. Thus, a prime objective of such assessments is to provide accurate information on verified performance (often called *M&V*) as rapidly as possible.

> **ASSESSMENT TERMINOLOGY**
>
> *Assessments*, in some cases within seconds, quantify and document actual demand flexibility performance of individual buildings with respect to defined metrics. In some contexts, such as for energy efficiency and demand response programs, these assessments are called *measurement and verification* (M&V). Assessments of time-varying retail rates are called *impact evaluations*. This term also refers broadly to a wide range of assessments of actual performance.
>
> By any name, assessments of actual performance are fundamental to advancing demand flexibility for buildings.

Building owners, operators, and occupants can use this information to understand whether demand flexibility strategies are providing the expected benefits, improve demand flexibility performance, and better control electricity bills. Utilities and centrally organized wholesale electricity markets use assessments as the basis for compensation under contracts (e.g., program incentives and market settlements). Assessments also are required for impact evaluations of time-varying retail rates. In addition, information that assessments provide is critical to evaluating program cost-effectiveness, supporting electricity system planning (including DER potential studies and comparing demand flexibility with other resources for meeting generation and T&D needs), and validating demand flexibility value.

State and local governments can use assessments to advance demand flexibility in support of their energy goals such as electricity system reliability and resilience, energy affordability, and integration of new generating resources and loads, as well as energy policies including hardening critical energy infrastructure, reducing peak demand, achieving renewable energy targets, improving air quality, and electrifying transportation and targeted building loads (Figure 1-3).[16]

Demand flexibility performance assessments also may inform progress with respect to energy programs like building energy benchmarking, strategic energy management, home energy rating systems, and building-level performance standards. Assessments can adopt integrated or holistic approaches that account for a jurisdiction's multiple energy goals and improve integration of multiple DER programs.

State and local governments have several opportunities to support demand flexibility assessments, including through the actions of PUCs, state energy offices, and municipal utilities. For example, as discussed further in Section 4, jurisdictions can:

- Encourage performance assessments (including for their own buildings), provide technical assistance to building owners and operators, and disseminate and report assessment results and the metrics, analysis methods, and implementation strategies used
- Adopt current best practices for demand flexibility assessments and support standardization efforts by utilities, regional grid operators, standards organizations, and federal initiatives
- Support advances in assessment practices, such as use of performance metrics, by implementing pilot programs for public buildings that demonstrate demand flexibility capabilities through documented assessments

[16] SEE Action Network 2020a.

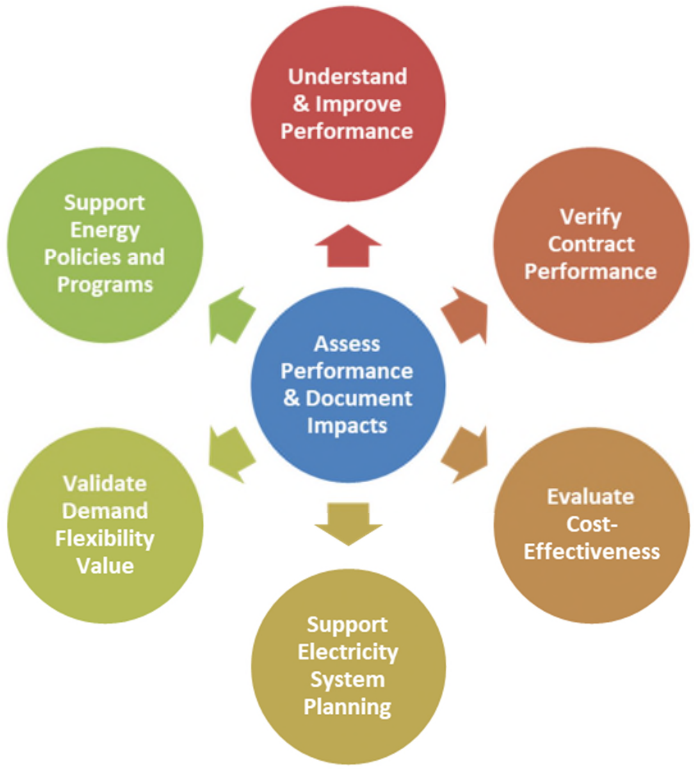


Figure 1-3. Multiple Objectives of Performance Assessments

Performance assessments of demand flexibility serve several purposes.

- Improve access to data from utility billing meters by adopting data access provisions, protocols, or standards
- Consider upgrading metering infrastructure and related communication standards and protocols through PUC guidance to regulated utilities and management or board guidance to publicly owned utilities and rural electric coops
- Improve coordination and communication across various utility programs and centrally organized wholesale electricity markets to enable participation by DERs, in concert with integrated demand-side management and GEB concepts for demand flexibility.

1.4. Assessment Phases

The assessment process consists of three phases: planning, implementing, and applying results (Figure 1-4). Applying assessment results includes using the information for the planning, implementation, and design use cases (Table 1-1). For example, using past assessment results can help assess future program cost-effectiveness. Section 2 of this report describes fundamental considerations for these phases.[17]

[17] Refer to SEE Action Network 2012, for more information.

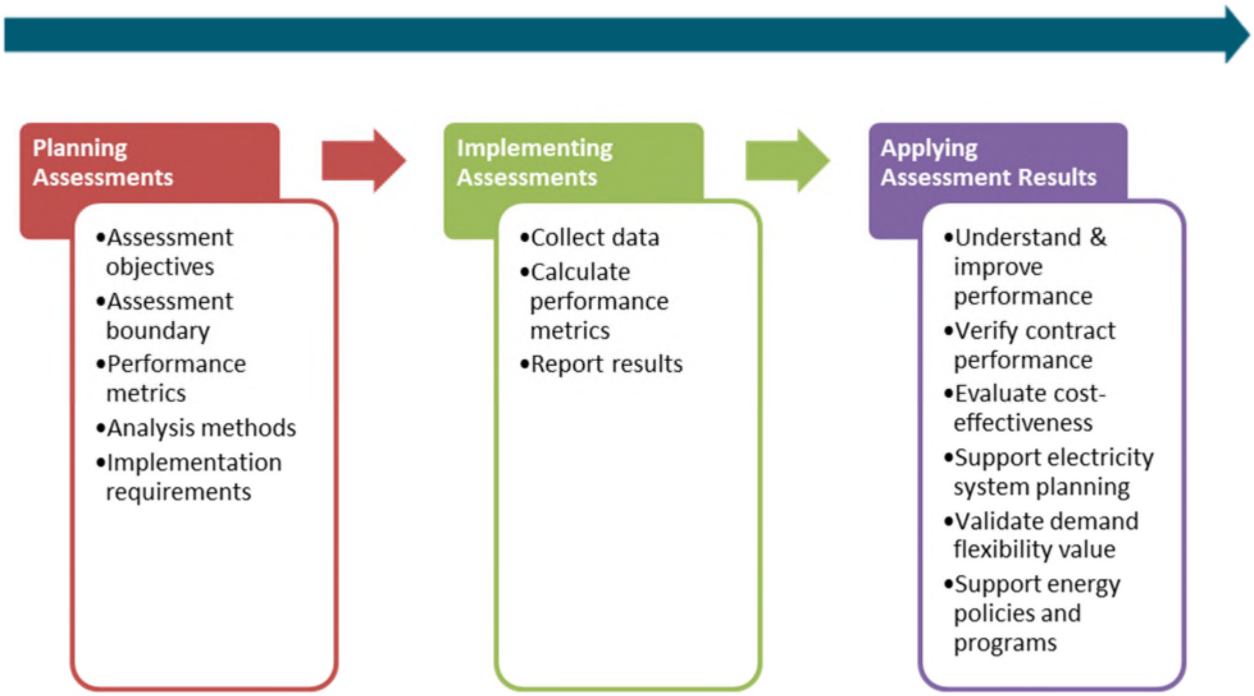


Figure 1-4. Demand Flexibility Assessment Phases

The three phases of assessments are planning, implementing, and applying results.

1.5. Current Demand Flexibility Assessment Approaches

Demand flexibility programs in the United States today focus on encouraging load shedding or shifting, either through *dispatchable* programs or time-varying retail rates.[18]

1.5.1. Current assessment approaches for dispatchable programs

Dispatchable programs provide financial incentives to consumers that change their electricity demand in response to events called by a utility, regional grid operator, or DER aggregator for reliability or economic reasons. These programs include a wide range of demand response approaches, such as direct utility control of air-conditioning using building

BASELINES AND COUNTERFACTUAL SCENARIOS

Baselines are typical conditions, including energy consumption and demand, which would have occurred without implementation of the subject demand flexibility activity. Baseline conditions are the counterfactual scenario, sometimes referred to as "business-as-usual." Baselines are used to calculate demand flexibility impacts. However, some demand flexibility assessments may not require baselines. In the future, new baseline approaches, or replacements for baselines, will need to be developed in the context of buildings providing continuous or near-continuous demand flexibility and multiple DERs operating in concert.

[18] For example, see Smart Electric Power Alliance and Navigant Consulting 2017. In this report, *dispatchable* means that a utility or regional grid operator controls the building's load and any generation and storage, or calls on the building operator to take a specific action, under a tariff or contract. *Non-dispatchable* means that the building owner or operator is deciding on their own to provide demand flexibility, typically to

thermostats and charging and discharging batteries.[19] Almost all current demand response programs use load shedding.

Assessments of dispatchable programs determine whether actual loads met program (contractual) performance obligations during demand response events. To comply with program requirements, a participant may need to demonstrate, through an accounting practice known as *settlement*, that during an event the building either simply took an action (e.g., a residence allowed the utility to cycle its air-conditioning system in response to a direct signal) or the building successfully modified its load by a specific amount relative to a baseline (e.g., shed a certain amount of load). For dispatchable shed events, M&V protocols and metering requirements are relatively well-established.[20]

1.5.2. Current assessment approaches for non-dispatchable utility tariffs

Load response under non-dispatchable utility tariffs activates solely at the discretion, or at least consent, of the building owner, operator, or occupant in response to price signals. Prices may be preestablished for each time period, such as with TOU rates with set peak and off-peak rates, or variable (not preestablished), such as with real-time pricing. Under critical peak pricing and peak time rebates, rates are fixed but event times vary within specified event windows. Consumer actions typically focus on load shedding or shifting and may be facilitated by technologies such as programmable communicating thermostats and "smart plugs." These tariffs also encourage behind-the-meter[21] generation, with or without energy storage.

EXAMPLE ASSESSMENT PRACTICES FOR DEMAND FLEXIBILITY PROGRAMS

Utilities in New England are developing programs for customer-sited DERs to reduce peak demand and ISO New England transmission charges. The programs allow customers to enroll eligible devices, including thermostats, water heaters, and batteries, for the utility's direct load control during peak events.

The utilities have developed simple assessment and settlement structures for residential battery systems. Eversource offers performance incentives based on the average battery discharge during events, with the $/kW incentive rate higher in the summer. Green Mountain Power uses a bill credit schedule that scales with the capacity that is committed for at least three hours. Green Mountain Power also has battery performance requirements. If batteries fail to provide capacity within +/-10% of committed capacity and program participants cannot correct the issue, the utility can remove them from the program. This verification of performance demonstrates how interoperable devices and automated analyses enable real-time or near real-time assessments, providing rapid feedback for settlements, performance improvement, and resource planning.

Sources: Green Mountain Power 2019, Mass Save 2019, and Trabish 2019.

1.5.3. Current assessment approaches for efficiency, generation, and load shed and shift

While existing M&V and impact evaluation procedures for energy efficiency have identified opportunities for improvement,[22] protocols are well established (Appendix C). Advanced M&V (see text box and Appendix C) uses

minimize overall energy costs. Demand flexibility also can be achieved without a financial transaction through voluntary calls from government agencies or grid operators to conserve or curtail load.

[19] See Cappers et al. 2011 for more examples of demand response program types.

[20] For example, see ConEdison 2020, as well as Section 2 and Appendix C.

[21] *Behind the meter* refers to DERs located on the building owner's side of the servicing utility's billing meter.

[22] Some opportunities for improvement relate to analytical methods and data access and collection that can support reduced M&V costs and increased reliability. There also is ongoing discussion on defining baselines and estimating attribution of impacts to efficiency programs.

ADVANCED M&V (M&V 2.0)

Advanced M&V (also called *M&V 2.0*) offers the potential to help realize the promise of buildings as grid assets in two important ways. First, it leverages the finer timescale of smart meter data. Second, automated analytics enable processing of larger volumes of data at higher speeds than ever before (Webster 2020). Perhaps as part of the increased application of advanced M&V, momentum is building for data standardization to support automated analysis of demand and energy impacts.

One example is Recurve Analytics' OpenEEMeter, which uses open-source software to analyze building energy savings (*OpenEEMeter* version 2.7.5 2019). The software leverages the established PRISM model for daily and monthly analyses and a model developed by Berkeley Lab for hourly analyses that can account for differences in building occupancy. The open-source approach facilitates third-party verification. The approach also supports a pay-for-performance model of efficiency in which aggregators of efficiency projects can be paid based on the performance of their projects, as assessed by M&V software (Best et al. 2019; https://www.lfenergy.org/projects/openeemeter/).

the same concepts as traditional M&V, but with more granular energy data, near real-time (e.g., within seconds) data access, and advanced analytics for more accurate and timely performance feedback.

M&V procedures also are well established for distributed generation projects. They are relatively simple, consisting of measuring generator output. When batteries are part of the demand flexibility strategy for a building, charging and discharging often are directly measured to assess performance.

For load shed and shift, the quantity of the load change is typically determined by comparing electricity consumption during an event with a counterfactual scenario or baseline level of consumption.[23] Advanced meters,[24] where deployed by utilities, facilitate collecting and reporting demand data that can be used for assessments. Whether using advanced meters or interconnected devices or BAS,[25] advanced M&V also can be applied for shed and shift assessments, as well as generation and storage.

The standard approach for assessing time-varying retail rates, particularly for residential participants, involves *impact evaluations* that compare electricity consumption patterns between large numbers of utility customers on time-varying rates with a control group of customers on fixed rates. These typically indicate demand flexibility performance, such as the amount of demand shedding during peak-period hours, over a year or more.

[23] Measured at the building's electricity service interconnection point between the building and the grid. Shed is by far the most common demand flexibility mode currently implemented for dispatchable programs. While sometimes used in combination with load shift or generate, performance metrics typically relate to shed.

[24] Advanced Metering Infrastructure (AMI) or Advanced Meters are defined by the U.S. Energy Information Agency as "Meters that measure and record usage data at a minimum, in hourly intervals and provide usage data at least daily to energy companies and may also provide data to consumers. Data are used for billing and other purposes. Advanced meters include basic hourly interval meters and extend to real-time meters with built-in two-way communication capable of recording and transmitting instantaneous data." See Form 861 Instructions for Schedule 6, Part D. https://www.eia.gov/survey/form/eia_861/instructions.pdf.

[25] *Building Automation System* (BAS) is a generic term for systems designed to control building operations and indoor climate. For the purposes of this report, they are assumed to include monitoring capabilities often associated with Energy Management and Information Systems (EMIS). EMIS is a broad family of tools and services to manage commercial building energy use. These technologies include, for example, energy information systems, equipment-specific fault detection and diagnostic systems, benchmarking and utility tracking tools, automated system optimization tools, and building automation systems. See https://betterbuildingssolutioncenter.energy.gov/alliance/technology-solution/energy-management-information-systems.

1.6. Future Demand Flexibility Assessment Approaches

To meet the full potential of GEBs to integrate multiple DERs, including energy efficiency, and multiple flexibility modes, new approaches will be required to assess demand flexibility performance (see Section 3)—for example, in the context of:

- Buildings providing continuous or near-continuous demand flexibility
- Engagement of multiple demand flexibility modes, in an integrated manner (Figure 1-5)
- Load modulation in sub-seconds to seconds to provide grid stability and balancing services[26]
- Increased use of combinations of DERs, such as on-site generation to charge batteries in concert with load shed and shift
- Integrated whole-building systems approaches to providing grid services as well as demand flexibility achieved at the individual end-use level or individual device level
- Managed EV charging.

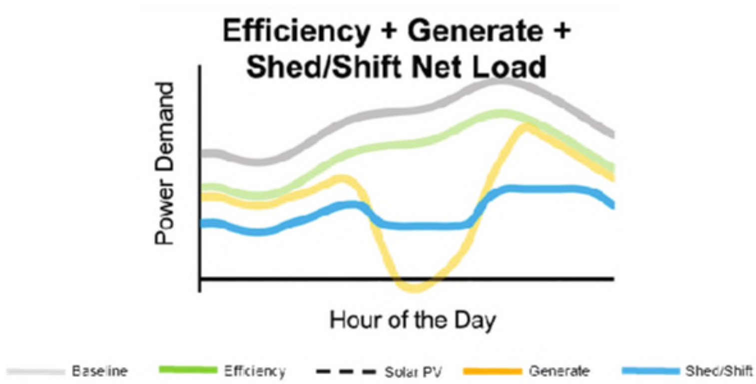


Figure 1-5. Demand Flexibility Modes in Combination: Shed, Shift, and Generate (with energy efficiency)[27]

Buildings are increasingly integrating multiple demand flexibility modes. This figure illustrates the changing demand over the course of a day of a single energy-efficient building with solar PV and load shed and shift.

Interviews with representatives of utilities, regional grid operators, DER aggregators, and state agencies indicate the need for protocols for assessments in these contexts (see Appendix A). However, these protocols have not been established. Addressing this lack of protocols for a future with more sophisticated demand flexibility services is addressed in Section 3 of this report.

[26] U.S. General Services Administration (GSA) 2019. Section 4.3 provides a brief discussion of grid services that modulation could provide and ways it could be implemented with respect to assessment development needs. Grid stability (frequency and voltage support) also can address building power quality issues. Seymour 2012.

[27] Neukomm et al. 2019.

2. Fundamental Considerations for Assessing Demand Flexibility Performance

This section discusses five fundamental considerations that define a demand flexibility performance assessment, with a primary focus on individual building assessments:

1. *Assessment objectives*–What information is the assessment intended to provide and how will the information be used?
2. *Assessment boundary*–At what level will performance be assessed—whole (individual) building or building system or equipment level, by DER, and/or by demand flexibility mode (e.g., load shed, load shift, modulate, or generate)? Or will the assessment boundary encompass multiple buildings, perhaps defined by their location (e.g., all buildings served by a single substation on the electric grid)?
3. *Performance metrics*–What metrics will be assessed and how will be they defined? What data will be required and at what temporal granularity (e.g., sub-seconds to hours for data measurement frequency) to calculate metric values?
4. *Analysis methods*–How will metrics be calculated and with what expectations for certainty? Will baselines be used and, if so, how will they be defined?
5. *Assessment implementation requirements*–What are the requirements with respect to data collection, privacy, cybersecurity, and reporting? What entities will conduct the assessments? What is the duration of (performance period for) assessments?

In particular, this section describes each of these considerations and current practices as they relate to demand flexibility for dispatchable programs and non-dispatchable time-varying retail rates. The next section discusses how to address several of these considerations for successful implementation of assessments in a future when buildings provide demand flexibility in increasingly sophisticated ways.

2.1. Assessment Objectives

The starting point for determining design and requirements for performance assessments is defining assessment objectives. Assessment objectives vary based on the audience for the assessment and how the derived information will be used—for example:

- *Building owners, operators, and occupants*, to stay below a certain level of demand during the billing period to avoid triggering a higher demand ratchet
- *State and local governments*, to meet goals for programs and policies and demonstrate compliance
- *Utilities, regional grid operators*, *and DER aggregators,* to determine compensation for buildings participating in programs and markets and to facilitate results reporting
- *Electricity system planners*, including utilities, regional transmission organizations/independent system operators (RTOs/ISOs), and state energy offices, that can use building performance data to inform forecasting and modeling for meeting resource and T&D system needs
- *Electricity system operators*, who can use assessments to inform real-time operation of systems
- *Energy service providers* that are developing and implementing demand flexibility services and products
- *Building designers (e.g., engineers)* that are responsible for new construction and retrofit designs of building systems.

To the extent practicable, state and local governments can encourage demand flexibility assessments that consider multiple key audiences where objectives are interrelated.[28]

As discussed in subsequent sections, defining objectives informs selection of an assessment boundary and performance metrics (and related measurement units of interest such as kW and kWh). In turn, the objectives, assessment boundary, and performance metrics define the other fundamental considerations discussed in this section—analysis methods and implementation requirements, including means for data collection.

Following are potential multiple objectives for performance assessments, most of which are important to one degree of another for all stakeholders (e.g., building owners and operators, aggregators, utilities, and regional grid operators):

- **Documenting demand impacts.** Rigorous assessments ensure that demand flexibility provided by a building is cost-effective and that impacts (e.g., demand reductions) are both real and sustained over time. For dispatchable programs with contractual performance requirements, assessments are oriented toward determining performance for the settlement process and resulting incentives (e.g., payments) or penalties for building owners or tenants and, if applicable, utility or other administrators of demand flexibility programs and third-party service providers (e.g., DER aggregators). For both dispatchable and non-dispatchable programs, documenting timing, quantity, and location of impacts also supports determining compliance with regulatory requirements for programs funded by the public or utility consumers, to ensure that funds are properly and effectively spent. One characteristic of documenting impacts is how quickly the assessment information is available. For most demand flexibility applications (and particularly for transactive, real-time applications), it is important to have quick feedback, essentially in real time so the grid operator can react, and within days to parties participating in a program or rate design (e.g., building operator and aggregator). This allows the parties to understand the effectiveness of their actions and financial impacts (e.g., incentives, bill impacts, and settlement payments).
- **Understanding and improving demand flexibility performance.** The role of assessments can go well beyond simply documenting impacts to actually improving existing and future programs and rate designs. Assessments reveal why program- or retail rate-induced impacts occurred and identify ways to improve demand flexibility activities. For example, assessments can provide information in real time to allow for as-needed course corrections and foster more effective programs and rate designs. In addition, assessments can support increased levels of investment in demand flexibility over the long-term for energy affordability, reliability, and resilience. The imperative for conducting performance assessments may be best described by Pearson's Law: "When performance is measured, performance improves. When performance is measured and reported back, the rate of improvement accelerates."
- **Supporting grid planning.** As demand flexibility is increasingly integrated into electricity system planning, assessments can support forecasting impacts at various geographic levels. Understanding and supporting the needs of forecasters and planners within state energy offices, utilities, and regional grid operators, as well as data formats and definitions, are important considerations for defining assessment objectives and metrics and reporting requirements.
- **Determining impacts on building functions.** Changing electricity demand profiles can impact the comfort and performance of building occupants and potentially affect operations and maintenance requirements of building systems. Assessments, perhaps combined with building commissioning and retro-commissioning activities,[29] may determine these impacts, and ways to minimize negative impacts and

[28] While not the subject of this report, market evaluation—another type of assessment—can be a valuable tool for increasing adoption of demand flexibility, to assess progress toward market transformation and other market-related objectives. SEE Action Network 2012 provides more information.
[29] See http://cx.lbl.gov/definition.html.

maximum positive ones, including earning financial incentives from utilities, regional grid operators, and DER aggregators.

Beyond these objectives, all directly related to demand flexibility, performance assessments can inform other building-related policy or program objectives, such as compliance with state or municipal building energy performance standards (see Section 1.3). It is important to take these other policy or program objectives into account when designing demand flexibility assessments so that the resulting information can be used effectively across all related applications.

2.2. Assessment Boundary

For demand flexibility assessments of individual buildings, the *assessment boundary* refers to whether demand flexibility is documented (measured or assessed) at the whole building level or for a specific piece of equipment, system, or DER within the building. Interviews with utilities, regional grid operators, and DER aggregators indicate that in many situations, the primary level at which demand flexibility performance should be assessed is at the whole building level—i.e., at the building energy meter (point of interconnection). This indicated preference for a primary assessment boundary at the whole-building level is based on two key considerations:

- If the assessment boundary is the whole building, a single whole-building meter that provides real-time or near real-time consumption information will provide all the data needed to assess the quantity and quality of demand flexibility services provided—at least if the measurement unit is kW or kWh. A key advantage is that the servicing utility or building owner already may have installed such a meter. If owner-installed, the meter may be stand-alone or integrated into a BAS.

- The building operator, or in some cases the utility, regional grid operator, or DER aggregator, adjusts end-use loads throughout the building to provide the desired demand flexibility. The bottom line for assessments is whether, at the point of service, the building provided the expected demand flexibility.

ASSESSMENT BOUNDARIES FOR INDIVIDUAL BUILDINGS AND AGGREGATIONS OF BUILDINGS

The primary focus of this report is assessing demand flexibility performance for an individual building. Going beyond a single building, another *assessment boundary* (the level at which performance is assessed) is defined by the aggregation of all buildings participating in a particular program or tariff. This is an important assessment boundary because, in most cases, demand flexibility must be aggregated across a large number of buildings to reach a magnitude sufficient to serve as a meaningful resource for electricity systems. For example, the assessment boundary for residential tariff impact evaluations using control groups is defined by all residences participating in the subject tariff.

However, in some situations it is advantageous or perhaps even required that the assessment boundary be defined by the equipment, system, DER, or flexibility mode. Following are examples of assessing performance contributions from such individual components. In the first example, demand flexibility is provided by a specific type of DER—in this case, a generator or storage system. In the second example, which can overlap with the first example, an assessment objective is to understand the contribution of each component—equipment, system, or DER—to the building's overall demand flexibility performance.

ASSESSING PERFORMANCE OF BATTERIES PROVIDING GRID SERVICES

In Hawaii, changes to net metering rules have driven adoption of batteries paired with distributed PV for residential utility customers. Starting in 2017, Sunrun, a provider of residential solar and batteries, installed batteries for all new solar installations. The company sizes solar and battery installations to the estimated energy usage of the home and potential solar production from the distributed PV system, aiming to provide as much of the home's electricity usage as possible from solar.

Each day, any solar production in excess of coincident usage charges the battery. Beginning in the afternoon, stored energy is discharged to cover home electricity usage in excess of solar production.

As part of a Grid Services Purchase Agreement between Open Access Technology International, Inc. (OATI) and Hawaiian Electric Company (HECO), Sunrun provides additional battery management to deliver grid value in two ways:

- Sunrun provides additional load reduction when the utility requests it during the evening peak period by discharging all available energy from the battery for up to four hours.
- PV system inverters provide fast-frequency services by responding autonomously to events when grid frequency goes below 59.7 Hertz (Hz). That typically can occur when there is a momentary imbalance between supply and demand. At these times, the inverter directs the battery to discharge for up to 30 minutes to rebalance grid frequency.

For control and assessment, OATI's Grid Services Delivery System, built from the company's webSmartEnergy software platform, aggregates resources from Sunrun (as well as other participating resources) for HECO. The system passes dispatch signals to a cloud-based system for communication to each device and shares telemetry data from the revenue grade meter that Sunrun embeds in each inverter, communicating via cellular chip or customer internet connection. Communication and settlement do not currently rely on utility AMI or building automation systems for performance data, but HECO is planning for utility AMI in the future.

Load shedding performance is assessed based on actual loads during events as compared to average demand during the event hours for the last 10 eligible (non-event) days. No baseline is used for fast-frequency response performance assessments. Settlements are based simply on whether the batteries discharged when requested and thus provided the desired frequency support.

OATI provides the M&V documentation for settlement. HECO reviews the documentation and audits the M&V as needed. HECO also has contracted with a third-party evaluation firm to provide overall review and assessment of the utility's demand flexibility programs.

For more information, see https://www.hawaiianelectric.com/products-and-services/demand-response/fast-demand-response.

Boundary definition also can include where on the grid the building is located and thus where demand flexibility services are provided. The location affects the value of the grid service.[30] For example, two buildings providing the same demand flexibility in terms of quantity, quality, and timing, but located in different places on the electricity system, most likely will not produce the same grid benefits (e.g., value of avoided energy and capacity). In addition, reducing demand in the wrong places at the wrong times could change power flows and compromise power quality and system reliability. Thus, monitoring demand flexibility by location as well as time period is essential, especially as demand flexibility grows.

[30] SEE Action Network 2020b.

2.2.1. Assessing Generator or Electricity Storage Performance

Generators such as distributed PV and electricity storage such as batteries can provide demand flexibility without relying on or impacting building functions or systems (e.g., lighting, heating, cooling, and ventilating). While generation and storage equipment can be operated independently of these building systems, overall building demand flexibility performance can be optimized when generation, storage, and control of a building's energy-consuming equipment and systems are coordinated.

Performance of distributed generation and energy storage systems are typically assessed without comparison to a baseline. Today (and until such time as distributed generation and storage become standard practices), the business-as-usual counterfactual is considered to be the subject building(s) operating without generation or storage. Any charging or discharging (in terms of kW and kWh output or input) can be direct indicators of grid services provided. Thus, demand flexibility performance of generator and storage systems can be independently monitored and analyzed. There is no need to define baselines as long as it is assumed that these systems would not be in place and operating in a manner that provides grid support without the program or rate design being assessed.

2.2.2. Assessing Individual Component Contributions to Overall Building Demand Flexibility

Multiple DERs may operate in multiple flexibility modes in buildings, affecting multiple pieces of energy-consuming equipment and systems. Particularly for large commercial buildings, a BAS responding to electricity prices or other factors may activate one or more DERs and flexibility modes to adjust power input and functional output (e.g., lighting, ventilation, cooling levels). Thus, in some cases, the objectives of the assessment require defining the assessment boundary at a more granular level than the whole building—say, for each piece of building equipment (e.g., motor, chiller, water heater), subsystem (e.g., heating, ventilating, and air-conditioning; lighting; or process), or DER. For example, there may be interest in understanding demand savings due to energy efficiency versus demand shedding, as may be the case with installation of more efficient air-conditioning equipment that includes a demand response-enabled control algorithm.

ASSESSMENT OBJECTIVES DRIVE METRICS

Some metrics for demand flexibility performance are likely to be the same regardless of assessment objectives (see Section 2.1). For example, most assessment objectives will require some indication of the amount of shed load, in kW, during a specified period of time. Additional metrics will be required to define any supplemental information needed to meet the assessment objective. For example, if the objective is to support electricity system planning or improve demand flexibility performance, additional information may be needed on how or why the impact occurred or did not occur.

Information from the assessment can be used to optimize performance of each piece of equipment, system, or DER leading to optimization of whole building performance, providing the most demand flexibility possible in a cost-effective manner that also is minimally disruptive to building occupants.[31] More granular assessment boundaries also can support the other three use cases for demand flexibility metrics—planning, operation, and design (see Table 1-1 in the Summary).

[31] For example, with granular data on the performance of each system, the assessment can identify the contribution of each system to the provision of demand flexibility.

2.3. Defining Metrics to Assess Demand Flexibility Performance[32]

Once objectives and boundary have been defined, the next and perhaps most critical step in designing an assessment is to determine the appropriate performance metrics. Metrics are numbers, or other forms of information (e.g., categorical values), that indicate the performance of the process of interest—in this case, demand flexibility. Metrics also point toward ways to improve the process. Metrics for demand flexibility may encompass four dimensions (see Figure 2-1):

- Quantity and timing of demand flexibility provided (e.g., amount of demand reduction during defined period in kW or kWh)
- Quality of demand flexibility provided (e.g., speed of achieving desired demand change or persistence of desired demand flexibility over long periods of time)
- Attribution of impacts to equipment, DER, flexibility mode, and/or building location (performance assessment boundary)[33]
    - Individual equipment (e.g., chillers) and systems (e.g., lighting)
    - Individual DERs
    - Individual flexibility mode
    - Location of impacts on the electricity grid
- Impacts on non-energy building services and occupants (e.g., comfort, health, and productivity).

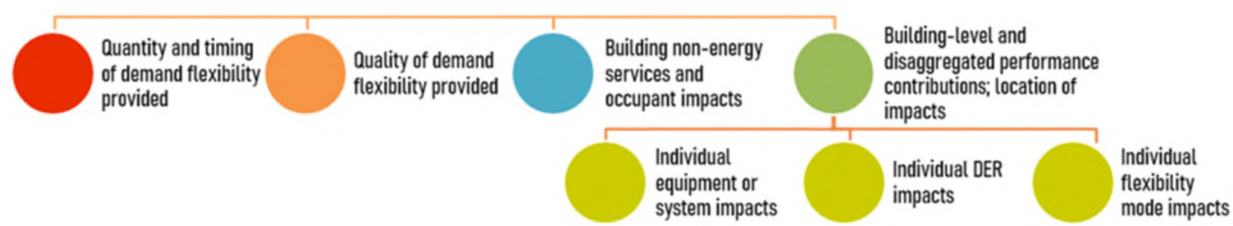


Figure 2-1. Organization of Demand Flexibility Metrics for Grid Services

Demand flexibility metrics for grid services may encompass four dimensions.

2.3.1. What metrics must be quantified and documented at the building level to assess demand flexibility performance for grid services?

For utilities and regional grid operators, the most important metrics for demand flexibility are associated with providing grid services.[34] Such metrics are determined by the flexibility mode, individually or in combination, and

[32] A DOE-sponsored project, "Framework & Method to Define Flexible Loads in Buildings - To Integrate as a Dynamic & Predictable Grid Resource" (see Liu 2019), is a collaboration between Berkeley Lab and Lawrence Livermore National Laboratory to develop metrics for demand flexibility modes. Initial research for this project on definitions for demand flexibility metrics informed this section of the report.

[33] For performance assessments of multiple buildings, the assessment boundary is expanded beyond the individual building, equipment, or system to encompass all buildings—or all of the specific equipment or systems of interest (e.g., all air conditioners in buildings on a specific time-varying retail tariff)—as well as the location of buildings on the electricity grid.

[34] Grid services include generation and delivery of electricity to consumers, as well as ancillary services. *Ancillary services* are "Those services necessary to support the transmission of electric power from seller to purchaser, given the obligations of control areas and transmitting utilities

grid services provided. Appendix D lists grid services that energy efficiency and each of the four flexibility modes can provide if they meet duration, load change, response time, event frequency, and other requirements. Generally, grid services provided by both energy efficiency and demand flexibility can be categorized as:

- *Reducing generation costs* by offsetting generation capacity investments, avoiding power plant fuel costs and operation and maintenance costs, or providing ancillary services such as frequency and voltage support and regulation and contingency reserves at lower cost
- *Reducing electricity delivery costs* by offsetting T&D capacity investments, increasing T&D equipment life, reducing equipment maintenance, or supporting voltage control.

System benefits and costs, the foundation of cost-effectiveness analyses, are fundamental metrics. Additionally, cost-effectiveness analyses should be aligned with a jurisdiction's applicable policy objectives.[35] Therefore, metrics that address non-grid impacts (e.g., community resilience, carbon reductions, economic development, and energy security) may be applicable.

Quantity and Timing Performance Metrics.

Metrics that address quantity (amounts) of grid services are defined by the flexibility mode and the grid service provided during a specific time period (and, depending on the grid service provided, at specific locations on the grid—see discussion of assessment boundaries). As one DER aggregator interviewed for this report noted, metrics for settlement for dispatchable programs are based on "what you get paid for."[36] Similarly, for non-dispatchable tariffs such as time-varying rates, what matters to the utility is the resulting demand (and demand load shapes) for participating customers. Thus, impacts are measured the same regardless of market structure (see text box).

Figure 2-2 outlines the path to defining quantity performance metrics for demand flexibility. Table 2-1 presents specific performance quantity metrics. Most of the flexibility modes and quantity metrics in the table are currently in use, at least to some extent. The primary exception is the modulate mode. It is not yet clear how this strategy will be implemented or what grid services it will provide, and thus what measurements and metrics will be required (see Section 3.3).

Figure 2-2. Process for Defining Demand Flexibility Performance Metrics

The path to establishing performance metrics is specific to the demand flexibility mode and grid service(s) provided.

within those control areas, to maintain reliable operations of the interconnected transmission system…." https://www.ferc.gov/industries-data/market-assessments/overview/glossary.

[35] NESP 2017 and NESP 2020. A new publication, *National Standard Practice Manual for Benefit-Cost Analysis of Distributed Energy Resources*, is forthcoming. See https://nationalefficiencyscreening.org.

[36] Utilities and regional grid operators need to consider what grid services are viable for program participants to provide and what metrics they can meet. Otherwise, programs may have poor participation or be very expensive. Robust stakeholder engagement, including aggregators, building owners, and customer groups, is important for designing programs.

Table 2-1. Quantity Metrics for Demand Flexibility[37]

Flexibility Mode	Grid Service	Measurement Units and Typical Metering for Whole Building Assessment	Examples of Performance Metrics
Shed Load	Generation energy or capacity, contingency reserves, T&D non-wires solutions	Power demand (kW) - For each load shed event: • average amount of demand over a period of time • baseline amount of demand over the same time period. Typical metering - Whole-building meters that record kWh and kW (e.g., at 5- to 15-minute intervals)	• Demand shed per event - Average demand (kW) reduction during a shed event or price-differentiated time window (e.g., for time-varying rates) measured against a baseline (such as historic demand during the same time of day, same day, and/or same weather) • Demand shed percentage - Demand shed per event divided by baseline average (or peak) building total demand (kW) during shed window • Normalized demand shed - Demand shed per event divided by the building's square footage • Demand shed per grid node (i.e., location of impacts)
Shift Load (moving load to time periods with lower electricity rates, also known as "Take")	Generation capacity, contingency reserves, T&D non-wires solutions, avoided renewable curtailment	Power demand (kW) - For each load shed event: • average amount of demand over a period of time • baseline amount of demand over the same time period. Energy (kWh) - Net total increase or decrease in building energy use during a defined period of time (including take and shed event windows) Typical metering - Whole-building meters recording kWh and kW (e.g., at 5- to 15-minute intervals)	• Demand shift per event - Average demand (kW) increase during a shift event or price-differentiated time window (e.g., in a retail tariff) measured against a baseline (such as historic demand during the same day types, with or without weather adjustments) • Normalized demand shift percentage - Demand shift/take per event divided by baseline average building total demand (kW) during shift/take window
Modulate Load	Frequency response and regulation, reactive power support (voltage support)	Power demand/supply (kW) and energy (kWh) Reactive power (kVAR) consumed/produced, upper and lower bounds—during each modulation event if providing kVAR support via control of reactive loads Grid voltage (V)/power factor and/or frequency if building is responding autonomously to grid conditions Typical metering - Whole-building meters or sub-meters that record kWh and kW. Depending on service provided, meters could measure kVAR, power factor, grid voltage, or grid frequency (e.g., near instantaneous intervals).	• Magnitude of demand response (kW) increase or decrease (or combined) for real-time service commitment (for frequency regulation) • Upper and lower bounds for kW increase and decrease (+kW, -kW) relative to a specified amount of demand (kW) for real-time service commitment (for frequency regulation) • The amount of time to full response (e.g., time to 90% of kW set point) • Whether controlled devices discharged power (e.g., batteries) or shut down (e.g., motors) in response to a defined load profile requested by grid operator
Generate	Generation energy or capacity, contingency reserves, T&D non-wires solutions, ramping, frequency response, regulation, voltage support	Power supplied (kW) Energy supplied (kWh) Typical metering - Meters that directly measure generator (or battery) output in kWh and kW (e.g., at 5- to 15-minute intervals)	• Average kW provided (generated or from storage) during an event • Average kWh provided (generated or from storage) during an event

[37] This table incorporates information from Liu 2019. Currently demand response can provide energy, ancillary services, load shifting, and emergency response. For details on measurements (and telemetry) needed, see a summary from the California Independent System Operator (CAISO): http://www.caiso.com/Documents/ParticipationComparison-ProxyDemand-DistributedEnergy-Storage.pdf. For performance metrics, see http://www.caiso.com/participate/Pages/Load/Default.aspx.

Performance metrics for grid services are determined by flexibility mode(s) and grid service(s).

Regardless of flexibility mode or grid service provided, primary metrics can be described with a limited number of measurement units, either in absolute terms or normalized (e.g., shed load by building floor area or peak building load):

- Two primary measurement units—demand (kW) and energy (kWh)
- Secondary measurement units (at the building/grid interface) that may be associated with modulation implementation are voltage (V), frequency (Hz), power factor, and reactive power (kVAR)
- Time period of interest—duration of demand flexibility impact (e.g., hours, minutes, seconds) and when the impact occurs (e.g., 4 p.m. to 7 p.m. weekdays).

For some flexibility modes and grid services, measurement units may translate directly to performance metrics (see Table 2-1). For other modes and grid services, measurement units must be compared to other quantities, usually a baseline, to understand the actual quantity of service (e.g., amount of load shed) provided by the building. For example, the amount of load shed during a specific time period is equal to the difference between actual load of the building and the load that would have occurred in the absence of participation in a utility demand flexibility program or time-varying retail rates. Defining and determining baselines for most metrics is a major component of the assessment process (see Section 1.4).

Measurement units will likely define performance metrics for demand flexibility in the following ways:

- *For load shed and load shift modes*, primary quantity metrics may be defined as quantified changes in the power draw (demand, kW) or consumption (energy, kWh) of a building as compared to a power draw or consumption baseline or

WILL DEMAND FLEXIBILITY METRICS VARY BY ELECTRICITY MARKET STRUCTURE?

Regional transmission organizations (RTOs) and independent system operators (ISOs) procure grid services differently than utilities do. Further, the value of demand flexibility may vary depending on electricity market structure—e.g., centrally organized wholesale electricity markets versus states without such markets and vertically integrated utilities—as well regulatory environment and availability of other electricity resources. However, energy, demand, and other impacts, and thus metrics, are quantified the same way.

Thus, as confirmed by experts interviewed for this report, what matters for defining metrics is whether the market or utility allows demand flexibility (and associated DERs) to provide particular grid services, such as generation capacity, non-wires solutions, and frequency regulation.

WILL DEMAND FLEXIBILITY METRICS VARY BY BUILDING TYPE?

No, according to experts interviewed for this report. While expectations for quantity and timing of demand flexibility vary by building type, metrics would likely be the same, based simply on flexibility modes and grid services provided (see Section 1.3). From a grid operator perspective, whether the building (i.e., the load providing the grid service) is a residence, office building, or industrial building is irrelevant.

However, only very large buildings with significant electric loads would be expected to have individual building impacts of interest to grid operators. Residential buildings, for example, would need to be aggregated for them to show significant impact for virtually any metric. Also, large commercial and industrial buildings have more sophisticated building automation and metering, providing enhanced capabilities for diagnostics and communications. Assessing demand flexibility performance beyond the whole-building level is more cost-effective for larger buildings.

load shapes—the counterfactual scenarios from which impacts are assessed.[38] However, it is possible that future assessments of demand flexibility will not require baselines (see Section 3.1).

- *For the generate mode*, metrics may simply require quantifying performance during specific time intervals and not require any baseline. For example, the metric may be defined as the amount of power delivered and equal to the power output of the generator (or battery) without consideration of a baseline. A conventional assumption is that the timing and amount of generation (or battery discharge) would not be adjusted to support grid needs in the absence of an incentive provided by a dispatchable program or non-dispatchable time-varying rates.
- *For the modulate mode*, the performance metric may simply be defined by whether the power at the building's interconnection point is within a defined range with respect to power draw—active or reactive (e.g., in response to a regulation signal sent by a grid operator). Or it may be defined in a manner similar to load shedding, as power draw relative to a baseline or specified amount of demand (see Section 3.3).

METRICS BEYOND PERFORMANCE ASSESSMENTS

Performance assessments support other use cases for demand flexibility metrics, including planning, operations, and design (see Table 2-1), by indicating if buildings have the ability to perform with respect to demand flexibility and to provide demand flexibility as a predictable resource.

Ability to perform
A state agency interviewed for this report identified, as information of interest, verification of a participant's *ability* to take demand flexibility actions. Assessments indicate if a building *has* performed, providing an indication of the *potential* of the building to perform—in other words, to provide grid services in the future.

Predictability
Discussions with RTOs and ISOs for this report emphasized the importance of predictability of load. As one individual put it, grid operators do not serve "savings," they serve load. Thus, grid operators need to understand whether loads participating in demand flexibility programs and time-varying retail rates will meet expected load profiles with respect to magnitude as well as shape.

These examples demonstrate that information of interest to program implementers and others should be determined in the program design stage, so that data collection and assessments meet program needs and support their effectiveness.

Quality Performance Metrics.

As demand flexibility becomes more commonly implemented and at more frequent intervals, and buildings provide more ancillary services, metrics beyond the simple quantity of any impact may become more important. A grid operator, for example, may require that resources meet committed quantities of reductions within a certain time frame or with a specified response time, duration of load change, or level of reliability or persistence such as percent of time available over a year[39] (see Section 3.2). Thus, additional metrics, for all flexibility modes, may include indicators of the quality of the demand flexibility provided by a building as a grid service. Persistence may be a particularly important metric for grid operators, including whether the building continues to provide demand

[38] Objectives and performance of efficiency and some demand flexibility modes may conflict with one another. For example, more efficient buildings may have less load to shed or shift.
[39] *Response time* or ramp rate is the amount of time from when a building asset receives a signal from the utility or regional grid operator until it responds to change the load. *Duration* is the length of time of the load change. *Persistence* is the duration of the demand flexibility service provided over an extended period of time, taking into account business turnover, early retirement of installed equipment, technical degradation factors, and other reasons services might be discontinued.

flexibility, in response to a dispatchable program or non-dispatchable time-varying rates, over extended periods of time and in response to frequent events.

Some of these quality metrics may actually border on quantity metrics, such as those that might be associated with compliance with a contractual requirement (as part of a dispatchable program), tailored to individual grid services or applications. These quality metrics indicate whether a building is providing a specific grid service, answering such questions as:

- What percentage of the expected load reduction, load shift, and electricity generation is the building providing for each event over a defined period time (e.g., a season)? In other words, what is the realization rate?
- Is the building persistently providing the expected grid service(s)—always, a minimum number of times, or on average? In other words, what is the compliance rate?[40]
- Does the change in building demand meet only a portion of the requested demand shed (e.g., percent of contractual load reduction) or meet the requested demand shed across all hours or just a portion of required hours?
- Is there an acceptable range of voltage and frequency support?

QUALITY AND QUALITY METRICS

The distinction between quantity and quality metrics is not absolute. This report refers to *quantity* metrics as those clearly associated with amounts (e.g., kW, kWh, and V). *Quality* metrics are additional indicators of the nature and value of demand flexibility. Examples include response time to a call for demand flexibility, ramp rate of load shift or shed during an event, whether demand stayed within certain parameters for voltage and reactive power, and sustainability (or persistence) of demand flexibility.

Examples of quality performance metrics are:

- Minimum kW shed provided by a building over a season
- Time for building load control to respond and reach the “average” kW reduction level, from the timestamp when a shed event starts.

Such quality metrics also can be used to analyze performance of the whole building, various demand flexibility modes, or specific DERs for the purpose of improving demand flexibility performance or for identifying future best opportunities for demand flexibility.

2.3.2. What metrics can be quantified and documented for building and occupant impacts?

While this report focuses on grid services, it is important for all stakeholders—particularly building owners, operators, and occupants—to understand impacts of demand flexibility on building functions. Table 2-2 provides some examples of such impact metrics, including four which have potential indirect economic impacts.[41] There

[40] Realization and compliance rates are similar, and perhaps overlapping, metrics. They indicate how well a building met a target value for a metric. Say the goal was to reduce demand by 100 kW between the hours of 4 p.m. and 6 p.m., for a total reduction of 200 kWh. If the reduction achieved during those hours was 180 kWh, the realization rate would be 90% (180/200). In addition, if the building reduced demand at a level of at least, say, 80% for 15 of 20 events a grid operator called during a year, then the compliance rate would be 75% (15/20).

[41] Examples of indirect impacts include additional staffing or maintenance needs and building comfort, which may adversely affect costs to operate the building as well as occupants’ willingness to pay.

also are direct economic impacts for building owners or occupants who are paid to participate in a dispatchable program or are on time-varying rates and see a net reduction in utility bills.

Table 2-2. Demand Flexibility Building and Occupant Metrics[42]

Impact Area	Measurement Units	Examples of Performance Metrics
Thermal Comfort	Zone temperatures (°F), predicted mean vote (PMV)*	• Average zone temperatures in areas affected by HVAC/plant demand flexibility strategy (°F) • Average zone PMV in areas affected by HVAC/plant demand flexibility strategy (unit-less)
Visual Comfort	Illuminance (lux)	• Average illuminance at desk level in areas affected by lighting or window demand flexibility strategy (lux)
Indoor Air Quality	CO_2 concentration, odors, particulate levels	• Average zone CO_2 concentration in areas affected by ventilation demand flexibility strategy (ppm)
Inconvenience Factor	Inconvenience factor	• Discrete levels determined by building operator or owner (e.g., "disruptive," "acceptable," "little impact") including amount of effort required to respond to price or event signals
Economic Impacts	Dollar value of participation	• Value of payments or savings for participation in program or on time-varying tariff (NPV $) • Costs for participation in program or retail rate (NPV $)

*Predicted Mean Vote (PMV) is a widely recognized metric for thermal comfort (+3: hot; +2: warm; +1: slightly warm; 0: neutral; –1: slightly cool; –2: cool; and –3: cold). According to ASHRAE 55-2017, the recommended PMV range for general comfort is between -0.5 and 0.5. NPV - net present value.

2.4. Analysis Methods

This section provides an overview of analysis methods for performance assessments. It discusses counterfactual scenarios and assessment reporting with a focus on existing assessment methods applicable to the way buildings currently provide demand flexibility. However, to meet the full potential of GEBs to integrate multiple DERs and flexibility modes, new analysis methods will be required to assess demand flexibility performance. These may be new approaches to defining baselines, or alternatives to baselines, for buildings providing continuous or near-continuous demand flexibility. Section 3 discusses development needs and opportunities.

What analysis methods are available for assessing demand flexibility?

There are two broad categories of current analysis methods:

- ***Direct measurement methods***—Methods involving direct measurement of performance metrics, thus not requiring a counterfactual scenario (i.e., a baseline). This involves measuring performance metrics in real time—for example, directly measuring energy output of a generator or amount of charging/discharging of a storage system or confirming the simple shutting off of electric water heaters in response to a frequency support request. Direct measurement-based analyses can be applied to individual buildings or a collection of buildings.
- ***Baseline methods***—Methods requiring definition of a counterfactual. Two subcategories are M&V and control group methods.
    - o *M&V methods* involve measuring in real time units of interest (e.g., kW, kWh) for a building and comparing those to a counterfactual scenario. M&V methods typically are used to assess individual buildings or collections of buildings participating in dispatchable demand response

[42] Adapted from Liu et al. 2019.

programs. Counterfactual scenarios are defined in a manner that represent historical load patterns, or what load would have been, at the subject building or buildings, without the applied demand flexibility modes. For example, direct measurements for a building's power draw during a weekday event from 2 p.m. to 4 p.m. can be compared with typical power draw measured for that building during recent, non-event weekdays from 2 p.m. to 4 p.m.

- *Control group methods* involve comparing measurements of interest (e.g., kW, kWh) for multiple buildings participating in a program or on a specific tariff ("treatment group") and for a similar group of nonparticipating buildings ("control group") during the same time period.[43] In practice, control group methods are only used for assessing collections of buildings participating in non-dispatchable tariffs. While they can be reliable and precise, control groups have disadvantages. First, the control group cannot participate in the program or tariff at the time of the study, limiting participation and thus impacts. Second, it can be difficult to find equivalent buildings that match the participant group, particularly for commercial buildings.

Figure 2-3 presents a hierarchy of assessment analysis methods. As Table 2-3 shows, application of analysis methods varies by demand flexibility mode, the services provided by each mode, and the type of program or tariff.[44] Table 2-3 shows methods that typically apply today to each demand flexibility mode—as well as to energy efficiency, for comparison—and that may be expected in the future as demand flexibility (particularly modulation) for grid services continues to develop.

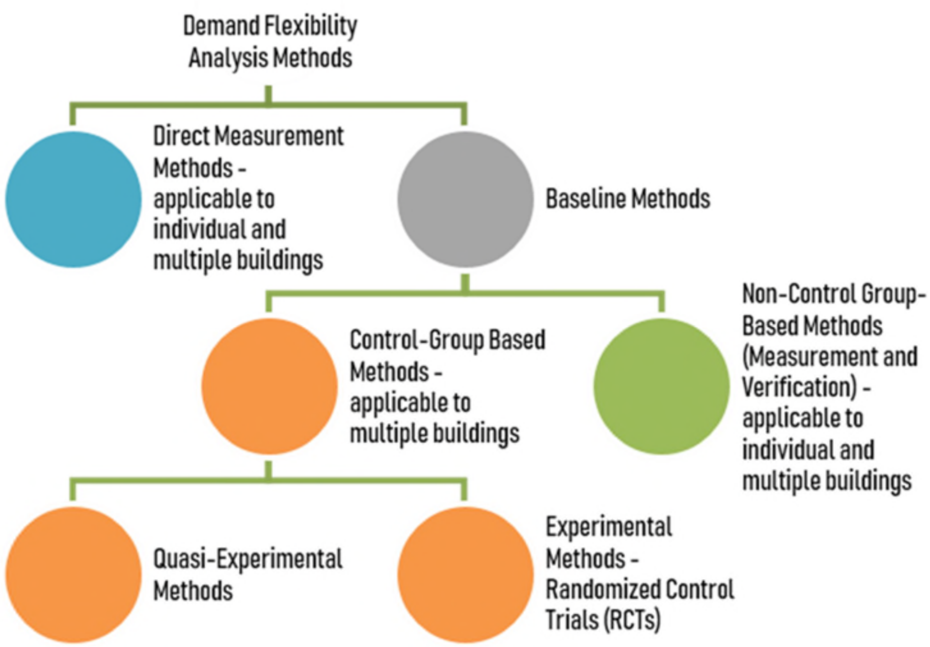


Figure 2-3. Hierarchy of Analysis Methods for Demand Flexibility Assessments

Analysis methods may use direct measurement or baseline methods. Table 2-3 uses the same color-coding as this figure. Because matching is feasible only in the context of multiple buildings, control group methods do not apply to individual buildings.

[43] *Randomized control trials* assign the study population to participant or nonparticipant groups based on chance (i.e., same opportunity to be in either group). In *quasi-experimental methods,* assignment of the control group is selected using a process that is not totally random.

[44] Appendix C provides more information and references for current assessment (EM&V) resources.

Table 2-3. Typical or Expected Applications of Analysis Methods for Energy Efficiency and Demand Flexibility

		Applications						
		Efficiency or Demand Flexibility Mode					Program Type	
Analysis Method		**Efficiency**	**Shed**	**Shift**	**Modulate**	**Generate**	**Dispatchable Programs (Event-based)**	**Non-Dispatchable Programs (e.g., tariffs for time-varying rates)**
Direct Measurement					■	■	■	■
Baseline Methods								
Measurement and Verification		■	■	■	■		■	■
Control Group Methods	Experimental Methods	■	■	■			■	■
	Quasi-Experimental Methods	■	■	■			■	■

The most common assessments today to quantify performance of load shedding and, to a lesser degree, load shifting are for:

- Event-based, dispatchable demand response–To determine settlements (payments)
- Non-dispatchable utility tariffs for time-varying retail rates (or other rate designs that encourage demand flexibility)–To determine customer response (impacts).

For event-based, dispatchable programs, the settlement assessment process is defined contractually. Current industry practice for settlement for load shed and shift involves M&V using baseline methods. Several RTOs, ISOs, and utilities have defined M&V requirements (see text box and Appendix C). Such event-based program assessments can be applied to an individual building or a collection of buildings. When used for a collection of buildings, the direct measurements or M&V results for each building are aggregated together. Conceptually, control group methods could be used for settlement processes. For example, the Electric Reliability Council of Texas (ERCOT) considers the use of control groups appropriate for settlement of aggregated, weather-sensitive, residential air-conditioning cycling and thermostat programs.[45] However, it is not standard practice due to difficulties identifying applicable control groups, particularly for commercial buildings. In addition, control groups usually require some potentially eligible buildings to not participate in a program or tariff (so they can populate the control group), and control group-based results tend to take longer and cost more than real-time measurements or M&V without control groups.

EXAMPLE ISO M&V REQUIREMENTS FOR DEMAND RESPONSE

ISO New England's Market Rule 1 describes basic rules for the operation of energy and capacity markets, including requirements for demand response resources. The rule defines units of participation (Demand Assets that represent single points of connection to the grid and aggregated Demand Resources)—and minimum demand reduction capacity. The rule also details the baseline methodology, including separate calculations for weekdays, Saturdays, Sundays/holidays, and day-of adjustments (which account for differences between baseline and metered demand just before an event).

Source: ISO New England 2019

[45] California ISO 2017; ERCOT 2019.

For non-dispatchable tariffs, assessments usually are designed to indicate performance over a year or more with respect to program objectives, such as consistent demand shedding during peak-period hours. Typically, demand patterns of a group of utility customers on a particular utility tariff that encourages load shifting and shedding are compared with demand patterns of similar customers that are on standard tariffs using a control group method. These assessments, known as *impact evaluations,*[46] indicate values for performance metrics collectively for all buildings in a program (see text box).

For energy efficiency, assessment practices have been well developed since the 1990s, based primarily on utility program impact evaluations and energy service performance contracts. Efficiency is a potential resource for all hours (measured in kWh) and also can provide energy capacity to serve peak demand (measured in kW).[47] Generally, efficiency can be assessed using any of the baseline methods in Table 3-3. Use of deemed savings also is common. Appendix C provides more information on these methods, known collectively as *evaluation measurement, and verification* (EM&V).

For generation, assessment practices are simple. They involve directly measuring the output of a generator or electricity storage system (when it is discharging).

For *modulation*, there are few examples of assessment practices. As Table 2-3 shows, in the future, modulation may be assessed with direct measurements (i.e., no baselines needed) or using M&V in a manner similar to how load shedding is assessed, but with shorter time frames.

EXAMPLE IMPACT EVALUATIONS FOR RETAIL RATES THAT ENCOURAGE DEMAND FLEXIBILITY

Assessments of time-of-use (TOU) and critical peak pricing (CPP) retail rates have employed experimental randomized control trials and quasi-experimental studies that use matched controls. These study designs allow investigation of the impact of tariff structure, rate level, and use of technologies that may enable demand reduction. For example, two randomized control trials evaluated time-varying tariffs for residential customers in combination with in-home displays that inform customers of prices, usage, and controls that allow customers to manage and schedule electricity-consuming devices:

- A study of a CPP tariff in Florida found 14% demand reductions on event days (Faruqui et al. 2017).
- A study in Connecticut found higher response from a CPP tariff than peak-time rebates (Faruqui et al. 2014).

To compare impact evaluation methods, Berkeley Lab used both randomized control trials and a 4-in-5 baseline method to evaluate the impact of TOU tariffs offered by Sacramento Municipal Utility District. The 4-in-5 method calculated an unadjusted baseline as the average consumption during event hours for the four days with the highest daily usage of the five most recent days, not including weekends, holidays, and previous event days. A comparison of demand reduction estimates from the two methods revealed that the 4-in-5 baselines led to underestimation of savings (Todd et al. 2019).

What are the counterfactual scenario considerations for assessing demand flexibility?

Grid-interactive demand flexibility involves an investment decision that is intended to result in grid and consumer benefits. As with other investments, it is difficult to assess what would have occurred, and what benefits would or would not have accrued, absent the investment. From this perspective, assessments are about answering the question, "What are *net impacts* of the decision to invest in demand flexibility?" Under current practices, any net changes in demand (or energy use) resulting from energy efficiency, load shed, and load shift cannot be directly measured. For example, the true impacts of demand shedding are the difference between the building electrical demand during an event and the demand the building would have otherwise required if the building operator did

[46] Goldberg and Agnew 2013.
[47] For example, see Demand Resources Working Group 2019.

not participate in a load shedding program or was not served on a time-varying tariff. In practice, we can never directly observe how much demand the building would have used had it not been in the program, because at any given time a household must either be in the program or not. Figure 2-4 illustrates this concept and use of the counterfactual scenario, or what is generally referred to as the *baseline.*

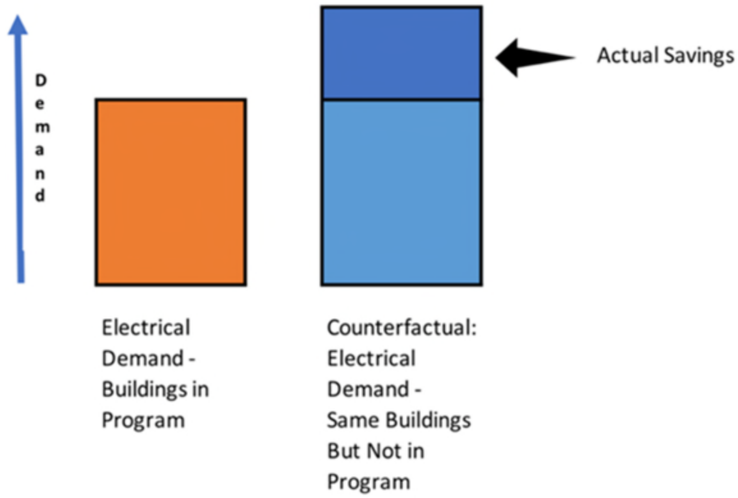


Figure 2-4. The Counterfactual (Baseline) Concept

Demand savings are the difference between participating buildings' electrical demand and the demand of the same buildings had they not been in the program during the same time period.

Two baseline methods are defined above—M&V and control group methods. With control group methods, whether using experimental or quasi-experimental approaches, the control group establishes the baseline of what demand (or other metrics of interest) would have been absent the provision of demand flexibility. This approach is the primary method used for assessing time-varying tariffs.

With M&V methods, the baseline is defined by historical electricity consumption or demand data for the building(s) providing demand flexibility.[48] Typically, recent days in which flexibility services did not occur provide a record of consumption that reflects normal operations. The counterfactual scenario can be derived from this record to estimate what consumption would have been had the building not provided a demand flexibility service using simple analysis (e.g., average consumption during the same time of day for the past 10 days) or complex tools (predictive estimating considering factors such as weather). M&V methods are most relevant for demand response programs. Counterfactual scenario assumptions for demand response assessments are day-matching and weather-matching, with or without regression analyses (see text box).

[48] If the building is new construction or significantly modified and therefore relevant historic data are not available, a building simulation model could be used to define the baseline. See Option D of the International Performance Measurement and Verification Protocol: https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp.

COUNTERFACTUAL SCENARIO (BASELINE) APPROACHES FOR DEMAND RESPONSE

M&V of dispatchable demand response services such as load shed typically depend on historical consumption data to define a baseline. Baselines estimate what electricity demand would have been in the absence of a demand response event. Two approaches are commonly used:

- *Day matching*–For the building being assessed, electricity use for a subset of days, usually in close proximity to the days leading up to an event, are identified and averaged to produce a baseline demand. For example, the assessment may determine the average demand during the subject hours (such as 4 p.m. to 6 p.m.) for the last 10 eligible (non-event) days.
- *Weather matching*–Weather-matching baselines are similar to day-matching baselines, except the baseline load profile is selected from non-event days with similar ambient temperature conditions and then calibrated with an in-day adjustment. Such in-day adjustments account for the difference between estimated baseline usage and actual usage before an event (e.g., 15 minutes). In general, weather matching tends to include a wider range of eligible baseline days, which are narrowed to ones with weather conditions closest to those observed during events. For example, the assessment may determine the average demand during the subject hours for five days with similar weather over the last three months. Regression analyses may be used for weather-matching adjustments.

California ISO's "Baseline Accuracy Work Group Proposal" (California ISO 2017) defines 14 day matching, and seven weather matching, baseline approaches and compares them to each other and a control-group baseline approach. Actual baseline methods that ISOs, RTOs, and utilities use vary by number of days and which days are selected, weather data used, rules for excluding data, treatment of prior event days, and same-day adjustments. See Appendix C for references to ISO and RTO guidance documents on baselines.

Is it important to assess attribution for demand flexibility?

Attribution estimates whether the actions of a participant in a program or tariff was influenced totally, not at all, or partially by that participation. This issue has been a major consideration in assessments of energy efficiency programs funded by utility customers, where the concept of free riders is often applied to determine net impacts.[49] In the context of this report, *free riders* are defined as participants in programs or tariffs who would have completely or partially replicated demand flexibility on their own and at the same time, or in the near future, in the absence of the program or tariff.

For DERs other than energy efficiency, the primary direct benefit to participants is payment for participation in a program or lower electricity bills, or both. It is not clear whether buildings would provide grid services using these DERs without programs or tariffs that financially reward building owners, operators, or occupants, although some participants may do so to reduce their carbon footprint or support a more reliable grid. Unlike energy efficiency, there are no obvious *direct* benefits to building owners or occupants to modify demand. Thus, at this time, free riders or attribution in general are not considered key elements of demand flexibility assessments. An important exception is buildings participating in more than one program at a time and eligibility to receive incentives from multiple programs (see Section 3.2). Attribution may become more important if more direct consumer benefits are identified for GEBs (for example, resilience benefits of on-site generation and energy storage for building occupants),[50] if more buildings provide demand flexibility as a societal good (e.g., to be considered green

[49] Consumer energy cost savings and other benefits from using less energy for the same service may be sufficient cause for people to invest in efficiency. There is a wide breadth of literature on this subject. For example, see Violette and Rathbun 2014.
[50] Installing such equipment for building resilience purposes may conflict with using that equipment to participate in a demand flexibility program.

buildings), or if building codes or utility interconnection agreements for generating DERs require demand flexibility.

What are the certainty expectations for assessments?

Certainty (or uncertainty) is a measure of the reliability of a value, such as a performance metric. All assessments involve a degree of uncertainty. Determining an acceptable level of uncertainty is an important aspect of defining an assessment. Uncertainty indicators refer to the amount or range of doubt surrounding a measured or calculated value. Thus, uncertainty is an overall measure of how well a calculated or measured value represents a true value. Uncertainty affects the reliability of commitments to deliver grid services and compensation for providing those services.[51]

EXAMPLE ISO PRECISION REQUIREMENTS FOR M&V

The "Measurement and Verification of On-Peak Demand Resources and Seasonal Peak Demand Resources" manual details ISO New England's M&V requirements for demand resources that bid into its Forward Capacity Market. Among the manual's key provisions, sampling must provide 10% relative precision with an 80% confidence interval. Statistically, *accuracy* is defined in terms of whether the confidence interval contains the true population parameter of interest (e.g., demand savings). The *precision* refers to the width of a confidence interval. The manual specifies how to calculate sample size to achieve the desired precision and confidence interval.

Source: ISO New England 2018.

All assessments seek to reliably determine energy and demand savings with reasonable certainty. However, the value of any reported performance metrics as a basis for decision-making, such as for settlement purposes, can be called into question if sources and level of uncertainty of reported metrics are not well understood and documented. Additional investment in the estimation process can reduce uncertainty. Still, trade-offs between assessment costs and reductions in uncertainty are inevitably required. Improved accuracy, and associated assessment costs, should be justified by the value of the improved information.

Sources of uncertainty in demand flexibility assessments include limited sensing and metering and limited availability of data at the desired resolution (e.g., data collection at specific timescales). Also important is uncertainty related to influences on building electricity demand, such as occupancy, weather, operational schedules, and parameters, which are needed to determine baselines. These are typically not measured.[52]

Using standard terminology, uncertainty of estimates for demand flexibility metrics is the result of two types of errors:

- *Systematic errors*—Errors that are subject to decisions and procedures associated with the assessment and which are not subject to chance; also called *bias.* Typical sources of systematic errors are measurement, data collection, and analysis. Another source of systematic errors is associated with selection of the baseline. This is applicable to any assessment method except for randomized control trials.
- *Random errors*—Errors that occur by chance. Random errors can come from changes in measurement units (e.g., demand, kW) due to unobserved influences (i.e., unobservable independent variables). Another source of random errors is sample selection to represent a population, such as a sample of buildings.

Because uncertainty arises from many different sources, it is usually difficult to identify and quantify the effect of *all* potential sources. Distinguishing between systematic and random sources is important because different procedures are required to identify and mitigate these two types of errors. Assessment protocols often identify

[51] U.S. DOE 2019a, 19.
[52] Ibid.

only uncertainty arising from random errors, because this source is usually the easiest to quantify using readily available confidence intervals and statistical significance tests to provide quantitative estimates of uncertainty caused by such errors (see text box on ISO New England M&V protocols for demand response).

Uncertainty attributable to systematic errors does not have a single comparable, quantitative measure. Rather, sources of systematic error (bias) are specific to individual studies, depending on equipment used, experience, and expertise of those conducting assessments, and assessment and data collection procedures employed. Assessing uncertainty from systematic sources requires that rigor of data collection, measurements, and analysis be addressed.

The result is that uncertainty is typically defined, calculated, and reported through objective analysis of random errors and subjective analysis of systematic errors.

2.5. Assessment Implementation Requirements

As Figure 2-5 shows, implementation of assessments involves information (data) collection, analysis, and reporting. This section covers design considerations for conducting these activities, with a focus on determining quantity and quality performance metrics related to grid services, including the following:

- How will the necessary data be collected (with consideration of privacy and cybersecurity)?
- What analysis tools will be used?
- What are the assessment reporting requirements, including who conducts the assessments?

Addressing these topics is a fundamental part of designing a successful assessment. Assessment planning can be influenced in an iterative manner by the practical realities of implementation.[53]

The first substantive activity for assessment implementation is collecting the necessary data. The next steps are analyzing, reporting, and using the information collected and analyzed. These activities can be overlapping and ongoing.

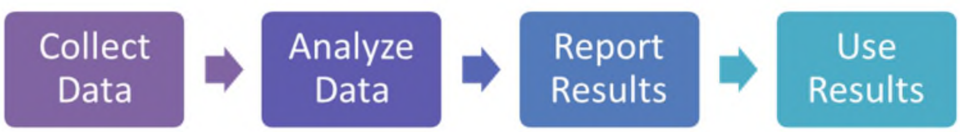


Figure 2-5. Assessment Implementation

Assessment implementation begins with collecting the necessary data. The information is then analyzed, reported, and used.

What are the means for collecting the necessary data?

In Section 2.3, the primary and secondary measurement units for assessing quantity and quality performance metrics are defined as demand (kW), energy (kWh), voltage (V), frequency (Hz), and reactive power (kVAR) and the time periods over which they are assessed. Access to these data at the whole-building level is typically through existing utility billing meters. Most commercial buildings have existing metering that can measure electricity

[53] A common saying is that in a plan, there is no difference between the plan and implementation, but in practice there is.

demand in sufficiently granular time increments (at least five-minute to hourly kWh data[54]) and communicate such information to various parties, including those who need the data for assessments. Some buildings have advanced metering through utility AMI deployment or use their own BAS to collect and convey data for analysis.[55]

However, there may be circumstances where (1) the building meter either does not provide real-time data at time intervals sufficient for assessing demand flexibility (e.g., residential meters that provide electricity consumption data at hourly or monthly billing intervals); (2) the performance metrics require additional data (e.g., kVAR or equipment/system level data); and/or (3) access to the utility meter data is problematic (due to delays in access to the data or privacy restrictions). Of particular importance is the ability to collect data at the required temporal granularity (e.g., sub-second to hourly), which varies by flexibility mode and grid service being assessed and the objective(s) of the assessment. In cases where the building's utility billing meter is not sufficient, additional metering would be required. This would likely consist of devices with their own sensing capability[56] or sub-metering connected to a BAS or directly to the data collection system of an aggregator, utility, or regional grid operator (see Section 3.2 for metering requirements for more sophisticated GEBs in the future). While this would provide massive amounts of additional data that can support assessment objectives, it also is potentially a burden for data management and verification of results at each device, DER, and/or building.

WHAT DATA ARE REQUIRED TO CONDUCT ASSESSMENTS?

Data, or measurement units, needed for assessment depend on the flexibility mode and grid service being assessed and the objective(s) of the assessment. For example:

- *Determining quantity and quality (including meeting compliance requirements) of flexible demand provided*—For most flexibility modes, demand (kW) measurements are the key data to be collected.
- *Understanding the "how and why" of performance*—Additional data are required, such as time-stamped control signals to and from building equipment and systems that are controlled as part of demand flexibility strategies (e.g., equipment on/off or load, HVAC system output, domestic hot water temperature, supply air temperatures), and current status.
- *Preparing baseline adjustments*—Weather or building occupancy data are required.
- *Understanding impacts of demand flexibility on building functions and occupants*—Additional information are required, such as indicators for comfort (e.g., interior temperatures), productivity, and impacts on utility costs.

Collecting data requires more than just meters. It also requires a data management system that takes into account:

- *Accessing data and information in a timely manner*–While data may have been collected by a utility meter or a BAS, access to the data by whoever is conducting the assessment may not be straightforward or timely. Thus, establishing a data conveyance mechanism is important. That includes interoperability and communication protocols—accepted rules and standards that allow communication and data-sharing

[54] While all electricity meters are able to provide kWh data, standard meters for residential and small nonresidential customers do not record demand (kW), and many meters do not measure reactive power and frequency. Thus, more advanced meters may be required for assessments of load modulation. See Section 3.3.
[55] Aggregators also may use the Green Button Alliance's Green Button Connect My Data for easy access to, and secure sharing of, utility-customer energy and water usage data. See https://www.greenbuttonalliance.org/about-cmd.
[56] Many DER technologies have integral metering capabilities, such as generators, electric (and thermal) storage systems, EV charging systems, and some Wi-Fi ready equipment such as water heaters and EV chargers. SEE Action Network 2020a includes example programs using these capabilities and interfaces.

between building automation equipment, aggregators, and utilities (e.g., Supervisory Control and Data Acquisition—SCADA—systems). For example, see the text box in Section 1, "Assessing Performance of Batteries Providing Grid Services."

- *Ensuring data quality*–Quality assessment processes and quality control actions are required to assure needed or desired accuracy to assess and document demand flexibility (e.g., meter accuracy and calibration standards).
- *Ensuring metering, equipment, systems and controls are private and cybersecure*–These components must maintain end-to-end data privacy, security and protection against unauthorized access while allowing secure communication of assessment information.[57]

An important aspect of information management is interoperability—the ability of two or more devices or systems to reliably and consistently exchange data. This is relevant for assessments so that any meters, building equipment, BAS, and other building systems can reliably and quickly provide data to those conducting assessments. DOE's *Quadrennial Technology Review 2015* identified interoperability as a "critical requirement for seamless integration of grid-connected devices, appliances, and building energy-management systems, without which grid modernization and further energy efficiency gains may be hindered."[58] See Section 3.2 for more information on interoperability and communications issues and opportunities.

What analysis tools will be used for analyses?

Conceptually, calculating performance metrics for demand flexibility is a simple process that can be completed with a relatively simple spreadsheet, database, or statistical software tool. However, handling large quantities of data and data cleaning (for example, to remove outliers) can require a significant amount of effort. This is particularly true for M&V methods if baselines are required. Numerous commercially available M&V tools are available for this purpose.[59]

In addition, "advanced M&V" (also called *automated M&V* or *M&V 2.0*) is particularly important for assessing demand flexibility. Advanced M&V has two key features: "(1) automated analytics that can provide ongoing, near-real-time savings estimates, and (2) increased data granularity in terms of frequency, volume, or end-use detail. Using [advanced] M&V methods that capture load shapes together with automated processing can determine savings in near-real time to provide stakeholders with more timely and detailed information."[60]

Today, greater data granularity for large numbers of customers, including from large AMI deployments, together with automated analytics can quickly assess performance of efficiency and demand response actions.[61] The concepts associated with advanced M&V also can be applied to future assessments of more sophisticated forms of demand flexibility, as discussed in Section 3.

What are the assessment reporting requirements?

Well-documented assessment results should be reported in a manner that fulfills the assessment objectives. For assessments solely focused on determining impacts, such as for settlement purposes, a contract defines reporting requirements, including time frames for reporting. These are important, as grid operators require moment to moment awareness of demand and available resources. For broader assessments, such as impact evaluations, additional information may be required, such as evaluations of how demand flexibility was achieved. The

[57] Navigant Consulting 2015; Mims et al. 2017; U.S. DOE 2016.
[58] U.S. DOE 2017a.
[59] Granderson and Fernandes 2017.
[60] Franconi et al. 2017, 3.
[61] Webster 2020.

frequency at which impacts need to be reported, and to what audiences (e.g., a regulatory body as well as a utility or regional grid operator), influence the scale of the effort and budget.

Who conducts the assessments?

Assessments (impact evaluations) of non-dispatchable programs, such as time-varying tariffs, usually are conducted by evaluation consultants[62] hired by a utility or, for regulated utilities, sometimes a PUC. For dispatchable, event-based programs, assessment data collection, analyses, and reporting may be conducted by the entity that is responsible for program implementation. When this occurs, reported impacts are "claimed."

The recipient of such reporting, such as a utility or regional grid operator that compensates the program implementer, may review the reports themselves or use an independent third-party expert to determine "verified" impacts. Independent experts are generally defined as having no financial stake in the assessment results (e.g., magnitude of savings). In addition, the organization, its contracts, and its business relationships should not create bias in favor of, or opposed to, the interests of the program administrator, implementers, program participants, or other stakeholders such as utility customers.

DISPUTE RESOLUTION

Disputes can occur between providers of demand flexibility, such as DER aggregators, and utilities or regional grid operators that pay for grid services through dispatchable programs. Under such "pay for performance" scenarios, disputes may arise about whether expected performance has occurred. Thus, implementation includes establishing dispute resolution processes. Such processes typically are part of contractual agreements between parties. They include provisions for access to data and analyses, as well as mediation or arbitration clauses, such as those suggested by the American Arbitration Association. See www.adr.org.

[62] Evaluation consulting firms tend to use engineers and econometricians—professionals who apply statistical and mathematical techniques to problem solving. For assessments of individual buildings, expertise in building systems engineering and experience with measurement protocols are important. For impact evaluations of aggregations of buildings, expertise in statistical analyses and experience with control-group based analyses are important.

3. Development Needs for Demand Flexibility Assessments

This section discusses development needs for demand flexibility assessments, now and in the future. As the table shows, virtually all aspects of demand flexibility performance assessments offer opportunities for improvement for determining impacts of each flexibility mode, now and in a future with buildings using multiple flexibility modes that are dispatched more frequently, potentially continuously, to provide multiple grid services.

Table 3-1 provides a high-level indication of the development status of energy efficiency and five fundamental elements of demand flexibility assessments discussed in the previous section:

1. Assessment objectives
2. Assessment boundaries
3. Performance metrics
4. Analysis methods
5. Assessment implementation requirements.

As the table shows, virtually all aspects of demand flexibility performance assessments offer opportunities for improvement for determining impacts of each flexibility mode, now and in a future with buildings using multiple flexibility modes that are dispatched more frequently, potentially continuously, to provide multiple grid services.[63]

Table 3-1. Development Needs for Energy Efficiency and Demand Flexibility Assessments

	Efficiency	Shed	Shift	Modulate	Generate*	Multiple Flexibility Modes, Dispatched Frequently
Assessment Objectives**	NA	NA	NA	NA	NA	NA
Assessment Boundary**	NA	NA	NA	NA	NA	NA
Performance Metrics	○	○	○	◑	○	◑
Analysis Methods	○	○	○	◑	○	◑
Baseline Definition	◑	◑	●	●	NA***	●
Assessment Implementation Requirements	◑	◑	◑	●	◑	◑

NA - not applicable as a development need
** Generate includes discharge from electricity storage (e.g., batteries).*
*** Objectives and assessment boundaries are critical to developing and implementing assessments. While they may be different in the future, they do not have development needs. Objectives are defined by the assessment sponsors; the assessment boundary flows from these objectives.*
**** Assumes baselines are not required for generation mode.*

○ *= limited development needs*

◑ *= low to moderate development needs*

● *= significant development needs*

[63] There also are opportunities to improve assessments for energy efficiency, another demand-side management strategy, even though such assessments have been conducted for at least 40 years. See the EM&V Appendix in Schwartz et al. 2017. For example, advances are underway for data collection and analysis. See Webster 2020.

Assessment practices addressing each of the five fundamental elements are in regular use today for basic applications of energy efficiency, generation (including storage charging and discharging), and load shedding. In particular, performance metrics and analysis methods are well established.

Thus, even for buildings that implement all of these modes, if the demand flexibility objective is to reduce or shed load, or generate electricity, as long as the assessment boundary is defined by the whole building, current M&V and impact evaluation practices are applicable and applied routinely. Even though it is not as common as load shedding, there is experience with assessing load shifting. Further, many of the practices associated with assessing load shedding can be applied to load shifting. For the modulation mode (Section 3.3), more experience and standard procedures are needed to address its fast response nature, what services it can provide, and how.

Building on As the table shows, virtually all aspects of demand flexibility performance assessments offer opportunities for improvement for determining impacts of each flexibility mode, now and in a future with buildings using multiple flexibility modes that are dispatched more frequently, potentially continuously, to provide multiple grid services.

Table 3-1, Table 3-2 describes three priority areas for development of procedures, or new standards or protocols, for improving consistency and credibility of future demand flexibility assessments:

- New baseline constructs
- Assessment implementation practices
- Requirements for modulation assessments, including sophisticated metering, communication and interoperability infrastructure at the building level, and definition of appropriate performance metrics.

A FUTURE VISION FOR DEMAND FLEXIBILITY PERFORMANCE ASSESSMENTS

Say a BAS for a large commercial building participating in a DER aggregator's day-ahead capacity market program receives a request on Tuesday morning to modify building operations for the next day in a manner that provides a specific electricity load profile. The requested load profile takes the form of specified hourly consumption levels based on electrical grid capacity needs and historic consumption and performance patterns of the building.

Once the building operator approves the request, load changes are implemented automatically. Using machine learning software and the results of actions taken for prior demand flexibility events, the BAS controls the building's multiple DERs to match the requested electricity load profile. The BAS uses real-time metered data to control a generator, charge or discharge a battery storage system, and adjust demand of the building's electricity-consuming systems (e.g., ventilation, cooling, and lighting), while keeping the building's performance within predefined parameters for occupant comfort and productivity.

During the demand flexibility event on Wednesday, the BAS uses its energy management information system attributes and the OpenADR standard to communicate individual DER performance data to the building operator and aggregator. These data include electricity consumption and generation data, as well as overall real-time electricity consumption of the building and calculated performance metrics according to the specifications defined in an agreement between the building owner and the aggregator. These might include, for example, compliance (or realization) rates and information on frequency and voltage of power delivered to the building's interconnection point. The BAS also transmits demand performance data, in real time, directly to the aggregator, which relays the information to the grid operator.

Early Thursday, the day following the demand flexibility event, the aggregator provides an electronic report to the building operator indicating whether the building provided the requested electricity load profile for each hour during the event, diagnostic information indicating how each DER contributed to the building's demand flexibility, and any recommendations for improved performance. The data also serve as the basis for transaction settlements between the aggregator and the building owner and grid operator.

This pattern is repeated multiple times each week.

While there are other opportunities for improvements for all types of assessments, they are not unique to GEBs and therefore are not a focus of this report.[64] At the same time, while improving implementation (e.g., metering, interoperability, communications, privacy, and cybersecurity) is applicable to all assessments, addressing this need is most critical for sophisticated demand flexibility assessments where data collection is required throughout the building, not just at the building's electricity interconnection point.

Table 3-2. Drivers of Priority Development Needs for Future Demand Flexibility Assessments

Drivers	Priority Development Needs
Increased use of shift and shed flexibility modes in combination creates overlapping baselines. Buildings providing multiple grid services, potentially in response to multiple programs offered by the servicing utility and/or regional grid operator, can create challenges for defining baselines. Increased number of events under dispatchable demand flexibility programs, including the potential for continuous demand flexibility, creates additional baseline challenges.	New baseline constructs (or replacements for baselines) will be needed in a future where one or more of these conditions occur.
Assessment objectives may expand to include understanding contributions to overall demand flexibility performance of individual equipment or a particular building system (e.g., water heating system), DER, or flexibility mode.	Improvements in assessment implementation practices (e.g., communication and cybersecurity protocols) and infrastructure (e.g., metering and advanced M&V) will be needed to address, among other issues, the large volumes of short interval data that result when many DERs and GEBs are providing demand flexibility at regular intervals and the need for analyzing impact information quickly to support grid operation.
Demand-side modulation is not a common practice today, although batteries are beginning to provide this service. Assessment practices will need to be established as this demand flexibility strategy is further developed with respect to services provided (e.g., frequency and voltage support), whether services are provided continuously or just in response to events, and whether implementation is in response to signals from the utility or regional grid operator or in response to direct measurement of grid conditions (e.g., voltage and frequency at the building/grid interface).	Assuming the modulation flexibility mode becomes more common—particularly with respect to being continuous, autonomous, and providing very fast response—metrics, measurement requirements, analysis protocols, and implementation protocols will need to be established.

3.1. New Counterfactual Scenario Constructs and Standards

Counterfactual scenarios for demand flexibility are estimates of demand that would have occurred absent dispatchable program events or price signals under a non-dispatchable tariff (e.g., time-varying retail rates). As introduced in Section 2, defining the counterfactual scenario is fundamental to some types of demand flexibility assessments.

[64] For example, while collection of building occupancy data may be needed for advanced baseline analyses, using building occupancy data already is a research area for assessing energy efficiency programs. Appendix C provides information on a range of M&V and impact evaluation topics specific to each flexibility mode.

Baseline practices have been established for efficiency, load shed, and load shift; generation assessments do not require a baseline (see Section 2.4). While there are opportunities for improving current practices, development needs for some demand flexibility modes do not restrict the ability to conduct informative assessments. As discussed in the following sections, however, some cases pose challenges for defining baselines for future demand flexibility applications:

- Shed and shift implemented together
- Multiple grid services provided by the same building under more than one dispatchable program
- Frequent provision of demand flexibility, including continuous demand flexibility.

3.1.1. Baselines for Combined Shed and Shift

With increasing amounts of variable renewable energy generation, and interest in minimizing its delivered cost and curtailment, the need for load shifting is growing. That includes increasing electricity consumption—sometimes called "take"—during times of high levels of electricity supplied by wind and solar, and when electricity prices are lower. Combining shedding load and shifting load may become an important strategy for demand flexibility in the future to address this need. For example, Bonneville Power Administration and Mason County Public Utility District tested demand flexibility for deferring a transmission system investment that otherwise would be needed to accommodate wind generation. Participating residential customers of the utility received communication hardware that delivered load-control information to enhanced water heating control systems. Using an automated demand response system, shedding and shifting of water heating loads helped maintain their coincidence with real-time wind generation (Figure 3-1).

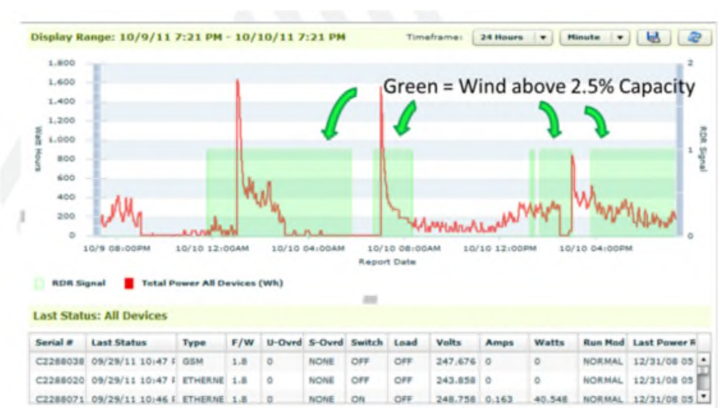


Figure 3-1. Demand Flexibility for Variable Renewable Energy Integration

When the red line (water heater load) is in the green shaded area, load is coincident with wind energy production.[65]

While progress has been made to define baselines for shed and shift working in conjunction, there is no consensus on the best baseline approach(es).[66] There are at least two challenges. The first is to define the baseline for load shedding when the building may be increasing load, through shifting, during times that are part of a typical baseline period—i.e., the hours or days before the shedding event. The second is to determine the net energy

[65] Sources: https://eta-publications.lbl.gov/sites/default/files/13._berry_and_miller_impacts_of_ders_on_transmission.pdf; https://www.bpa.gov/EE/Technology/demand-response/Pages/residential_demand_response.aspx.
[66] California ISO 2017.

impact of load shedding and shifting—i.e., whether total electricity consumption increased or decreased. For example, precooling a building to move load earlier in a day almost always requires more electricity overall, which could affect the amount of net load shed. Additional research, such as comparing demand reductions determined through randomized-control trials with other assessment approaches using various approaches for defining baselines, may help clarify how best to set baselines for assessments of dispatchable shed and shift programs.[67]

3.1.2. Baselines for buildings participating in multiple grid services programs

The potential for GEBs to provide multiple grid services introduces further complications for baselines. This may occur in areas where retail demand response programs administered by a utility or DER aggregators operate alongside wholesale market programs administered by an ISO or RTO. In these cases, a building's participation in one program may confound the baseline period for another program. For example, during one week a building may reduce demand to alleviate distribution constraints through a utility program and a week later respond to an RTO's call to reduce demand because of a generation capacity constraint. If the RTO is using a day-matching or weather-matching method, the day that included the first week's event may suppress the baseline.

This overlap in activity reveals two additional, interrelated issues for demand flexibility baselines. First, absent any data-sharing protocols, the RTO may not be aware of the building's demand reduction due to participation in a retail program.[68] Second, it is not clear whether participation in one program should represent normal operations for a second program. Utilities, regional grid operators, regulators, and providers of demand flexibility services can overcome these issues by developing data-sharing protocols, with stakeholder engagement, and determining whether baselines should account for events called by other entities.

3.1.3. Baselines for continuous demand flexibility

As buildings implement demand flexibility modes more frequently, additional baseline issues will emerge. For example, there will be fewer days without implementation of demand flexibility and thus fewer days will qualify for baseline calculations—those able to represent "typical" days with respect to electricity demand. Baseline-dependent M&V methods will fail when an insufficient number of days meet baseline criteria. Similarly, as more buildings provide demand flexibility services, the number of buildings that can be considered for inclusion in a control group for an impact evaluation will decrease. In both cases, there could be a future without "typical" days or buildings against which to compare the demand flexibility performance of buildings participating in an event.

Our interviews for this report with industry experts (see Appendix A) identified potential paths for demand flexibility assessments that bypass these baseline issues. Suggested approaches include the following:

- *Greater reliance on non-dispatchable, price-driven regimes (e.g., real-time pricing)*—These include utility tariffs for time-varying pricing that prompt building owners, operators, and occupants to consistently modify demand to provide grid services. There are no incentive payments or penalties, and thus no need for settlement M&V. Assessments would continue to use impact evaluations with control groups that define the counterfactual scenario. While it may be challenging to find appropriate control groups, certainty as to how representative control groups need to be and overall reliability of impact calculations would likely be less stringent given the lack of settlement requirements.
- *Development of new performance metrics*—Utilities and grid operators need predictability to serve loads. Thus, they need to know if actual loads met expectations for demand flexibility in terms of magnitude as

---

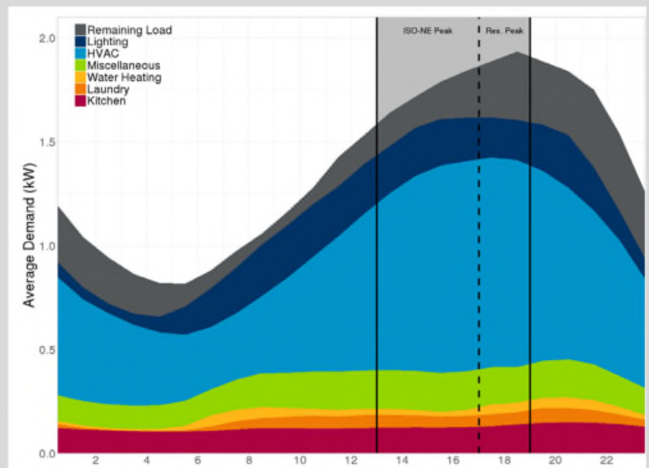
[67] Research needs for residential and commercial buildings are likely to be different with respect to the assessment of a combined load shed and load shift strategy.
[68] National Grid and the Energy Efficiency Advisory Council 2019.

well as load shape (see text box). In a future with continuous demand flexibility, buildings could commit to provide the desired load shape with maximum and minimum quantities of electrical demand for each hour (or perhaps each half or quarter hour). The performance metric is whether the desired load shape and magnitude matched what the building provided. No baseline is needed. Other metrics could include:[69]

  - Shaping options, where total energy savings for a period could be allocated or "shaped" among all hours in that period based on load levels[70] (see text box)
  - Annual or monthly peak demand—e.g., the highest 15-minute demand in a year or month, perhaps similar to ratchet provisions in some electricity tariffs[71]
  - Load factor—e.g., the annual average demand divided by the annual peak demand
  - Peak ratio—e.g., the 5% highest demand in a year divided by the annual peak demand.

- *Assessment of generation and storage discharge from distributed resources*—Assessments for buildings with distributed generation assess generation capacity and energy provided in the same manner as is current practice for evaluating distributed PV, batteries, or other generation and storage resources. No baseline would be required. The amount of power and energy provided to the grid varies based on *net* building generation (storage discharge), a function of generator (battery) output, building demand, and demand flexibility provided by all DERs.

LOAD SHAPES



A load shape is a curve that represents load as a function of time. Load shapes contain information on how electricity use changes over the day, as a composite of end uses such as lights, appliances, and heating, ventilating, and air conditioning systems. Load shape analysis is commonly used by building owners, operators, and energy managers to analyze energy consumption. Researchers have developed methods to derive these curves using historical electric meter data.

Source: Luo et al. 2017. Figure from Navigant 2018.

Establishing practical and accepted approaches to address these issues will require RD&D.

[69] See Liu 2019 for additional information on these and other metrics.
[70] The concept of shaping has been proposed by the California PUC based on the concept that "Where individual Energy Efficiency measures produce savings during the operating hours of the affected end-use application…, program administrators believe a portfolio of Energy Efficiency measures affecting a cross-section of end-use applications have been demonstrated to produce savings that are a function of the end-use load shape—i.e., that the amount of savings produced by a portfolio of Energy Efficiency measures is directly proportional to the amount of energy consumption." Demand Resources Working Group 2019, 6.
[71] Ratchets calculate the customer's electricity demand charge for the billing period based on their maximum demand, or a percentage of the maximum demand, for a period such as the prior 12 months.

3.2. Improving Assessment Implementation

A future with increasing levels of demand flexibility provided by sophisticated GEBs using multiple DERs and demand flexibility modes may lead to greater interest in assessments at the DER, device, or system level. Understanding the contributions of each DER, and for each major piece of building equipment or system, is important for determining contributions from each component of a GEB. This understanding also is important for predicting, planning, and operating GEBs in the most effective (including cost-effective) manner. The implication is that assessment boundaries may be more granular than those used for assessing performance at the whole-building level (i.e., using the electricity meter for the building).

More GEBs, more granular assessment boundaries, and shorter time frames (e.g., seconds) for demand flexibility actions all indicate that there could be orders of magnitude more data for analysis, and possible use in settlements, than available with today's demand flexibility programs. The large volume of data likely will require upgrades to aggregator, utility, and grid operator information technologies and systems. Data reported to a utility or regional grid operator for settlement could be aggregated (one data stream representing an aggregation of many small resources located in a similar geographical location). But such aggregations would make independent verification of meter data more complicated.

Time frames for reporting performance of demand flexibility resources also are a significant factor. Grid operations require moment-to-moment situational awareness, including awareness of what resources are available and what these resources are capable of doing if used. Implementing metering and telemetry systems to obtain real-time information on the status of thousands of DERs' availability, deployment, and performance could be a financial burden compared to monitoring the status of a single large generator.

These issues lead to additional assessment implementation challenges (also see Section 2.5) related to data collection, interoperability and communications, privacy and cybersecurity, and the role of third-party evaluators. These challenges are discussed next. While investments in metering and other infrastructure to address assessment challenges have costs, such infrastructure also is valuable for implementation of demand flexibility resources. Thus, the added costs for such investments may be incremental and small.

3.2.1. Incremental metering

Future assessments of demand flexibility, for sophisticated GEBs, may require more than just whole-building demand data recorded at intervals of five minutes or longer. Depending on the type of response required from the GEB, additional data requirements may include:

- Data collection with greater time resolution, down to seconds or sub-seconds, recorded by sub-meters for equipment providing the subject grid services
- Measurements of power consumption beyond just demand (kW), including voltage, reactive power, and perhaps frequency
- Weather and occupancy data for improving baseline determinations
- Building service and occupant impact data (see Table 2-2), to understand the impacts of providing demand flexibility on building functions and performance (e.g., comfort and air quality)
- Individual equipment and system power draw, performance, and status information to allow assessments at the DER, building equipment, building system, or flexibility mode level
- Potentially, real-time or near real-time monitoring.

Accessing such data requires reliable and accurate measurements and recording and transmitting data. Billing meters that the servicing utility locates at the electricity interconnection point(s) to a building can provide some of the data. But to meet the data needs listed above will require devices that can record and communicate their own usage,[72] sensors combined with sub-metering within a building, advanced billing meters, or a combination of these approaches.[73]

For equipment or systems that do not have the ability to monitor and report their own energy use, or if the existing BAS does not provide the needed data, upgrades likely will be needed to the EMIS capabilities of a BAS, and/or additional sensors and sub-meters within the building may be needed to measure and record in near real-time at very short time intervals. Such sub-metering may provide its own telemetry[74] or be connected to a BAS or an advanced building meter designed to collect such information.[75] Even residential and small commercial buildings may have some type of BAS or communications systems built into DER components (see text box on Hawaiian Electric programs, earlier in this report). Thus, the BAS is a likely means for data collection as its role already is defined as controlling and monitoring HVAC, lighting, and other building systems. By design, a BAS receives information from a range of sensors to provide information such as building, equipment, or system power consumption and outdoor and indoor air temperatures. A BAS can communicate this information to building operators and those needing such information for assessments.

3.2.2. Interoperability[76] and communications

Access to energy and demand data is necessary to enable interoperability of building equipment and systems and DERs, to allow DERs to provide grid services, and for assessing demand flexibility performance. Utilities, regional grid operators, and service providers need the data for settlements. Data are often building-level, but may include end-use (e.g., equipment and systems) level data as technologies and grid services evolve. Building owners and managers need access to equipment and system performance and energy consumption data to make informed decisions about demand-reduction strategies and operations. Incomplete collection and communication of data are impediments to assessments.

Communications protocols that support interoperability establish standard methods for making data more accessible and automating their use. Well-established communications protocols relevant to flexible buildings include:[77]

- *OpenADR2.0*—This standard enables communication among grid operators, demand service providers, utilities, and building occupants to support automated demand response. BASs and end uses such as EV chargers, HVAC systems, and lighting can be programmed to respond to signals. These signals may contain directions to reduce load for a specified duration and can convey price and event information. Use of OpenADR2.0 is well established for demand response programs.[78]
- *BACNet*—This international standard communications protocol for BASs was developed under the auspices of the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). BACNet defines how a BAS and devices communicate basic functions, such as operating schedules, and

[72] Nordham et al. 2019.
[73] An example of a program that supports sensors, meters, and other equipment, along with data analytics and information services in buildings, is NYSERDA's Real Time Energy Management Program, https://www.nyserda.ny.gov/All-Programs/Programs/Real-Time-Energy-Management.
[74] Telemetry involves collecting data (e.g., demand measurements) at one location and transmitting the information to monitoring equipment at another location.
[75] For example, see https://www.itron.com/na/solutions/product-catalog/openway-riva-centron-meter.
[76] For a description of interoperability in the context of smart grid, see IEEE 2011.
[77] The next two lists are not an exhaustive treatment of the protocols associated with interoperability and communications that are used, or may be used, in the context of assessing demand flexibility performance. Instead, the lists provide an indication of the range of such protocols.
[78] OpenADR 2.0 Profile Specification B Profile 2013.

report temperature and device status.[79] However, BACnet is a standard for functional controls and does not cover grid communication. Thus, another interface is required for transmittal of data to and from the grid.

- *CTA 2045*—This standard defines specifications for sockets on appliances that can vary their electrical input—e.g., shed load. Utilities or other entities can provide communications modules that plug into these sockets and connect the device for use in utility programs. Pilots have demonstrated the standard and associated technology, but costs have limited its implementation.[80] Variation in functionality and communications protocols across manufacturers makes product integration challenging. Utilities also may have to go through manufacturers to access device-level data,[81] which creates barriers for assessments. Efforts are underway to improve and modernize the standard.
- *IEEE 2030.5*—Similar to OpenADR2.0, this standard supports communication between utilities or regional grid operators and DERs, including demand response, EVs, and distributed PV systems. The standard has not been used widely for demand response.[82] California's Rule 21 requires use of this standard for distributed PV smart inverters. Utility demonstrations suggest that the functionality of IEEE 2030.5 is insufficient for DERs bidding into wholesale electricity markets.[83]

The above protocols establish mechanisms for exchanging information within a building or between devices/building(s) and an external party such as a utility or DER aggregator. How information is structured also is important for interoperability and conducting performance assessments. Following are examples of guidance related to information carried by communications systems:

- *Brick* and *Haystack*—Schemas for control systems are often site- or vendor-specific. That means any analysis of the system requires a customized approach. Brick aims to provide a consistent framework for describing equipment and sensors in a building and the relationship among them.[84] Haystack is an effort to standardize taxonomy of building automation data to enable scalable analytical tools that can quickly inform building managers of the potential for, and impacts of, demand flexibility.[85] ASHRAE is currently working to propose a standard that incorporates both Brick and Haystack into BACNet.[86]
- *Facility Smart Grid Information Model*—This model supports end-use load management in response to signals and relaying information about load to utilities and regional grid operators by creating a standardized information model that covers on-site generation, demand response, storage, and demand forecasting.[87] The standard maps to OpenADR2.0 and requires user testing.[88]
- LonMark—The goal of LonMark is consistent labeling of device functions ("functional profiles") so that devices can be managed regardless of manufacturer.[89]

[79] Newman 1997.
[80] Narang et al., forthcoming; Dayem 2018.
[81] Dayem 2018.
[82] Narang et al., forthcoming.
[83] AEIC et al. 2018.
[84] Balaji 2018.
[85] Prairie et al. 2016.
[86] See https://www.ashrae.org/about/news/2018/ashrae-s-bacnet-committee-project-haystack-and-brick-schema-collaborating-to-provide-unified-data-semantic-modeling-solution.
[87] Bushby 2011.
[88] Narang et al., forthcoming
[89] See http://www.lonmark.org/.

- *Open Building Information Exchange* (oBIX)—oBIX aims to standardize connection of controls and end-use data to the cloud, bypassing wired connections for a BAS.[90]

A major gap in interoperability (and communications) is how to standardize the translation of grid service requirements into automated building responses. While there are ways to communicate demand response needs for shedding load, and standards for interconnection and interoperability of DERs (IEEE 2018), today there are no standards for how grid operators define and quantify other services that GEBs can provide, such as reactive power and voltage support. There also is no standard for mapping grid services to automated operational changes for connected loads. Such definitions and mapping standards form the basis for communication and interoperability across the wide range of building systems (and equipment) and the grid. The Grid Modernization Laboratory Consortium (GMLC) has identified this as a priority for incorporation into OpenADR2.0 and also introduced the idea of the Energy Services Interface as a means to reconcile these gaps in translating grid needs into automated building action.[91]

Standardized mapping of grid services to automated building operations is important for demand flexibility assessments. This process connects the desired outcome (the grid service) to devices or systems that provide that service (e.g., HVAC, battery) and thus the connections required between the grid and GEB systems. With mapping, assessments could be more granular and evaluate the relative contribution of individual end uses to each grid service. Building managers would be able to better plan their operations and optimize their contribution to grid services. Standardized mapping also could facilitate development of open analytical approaches. Open-source analytical approaches may increase confidence in assessments. Figure 3-2 is an example framework that maps connections (interfaces) between the grid, master controller (BAS), and individual controllable components in a building. The figure illustrates the importance of interoperability—the ability to exchange data between the electricity grid, controls, and devices.

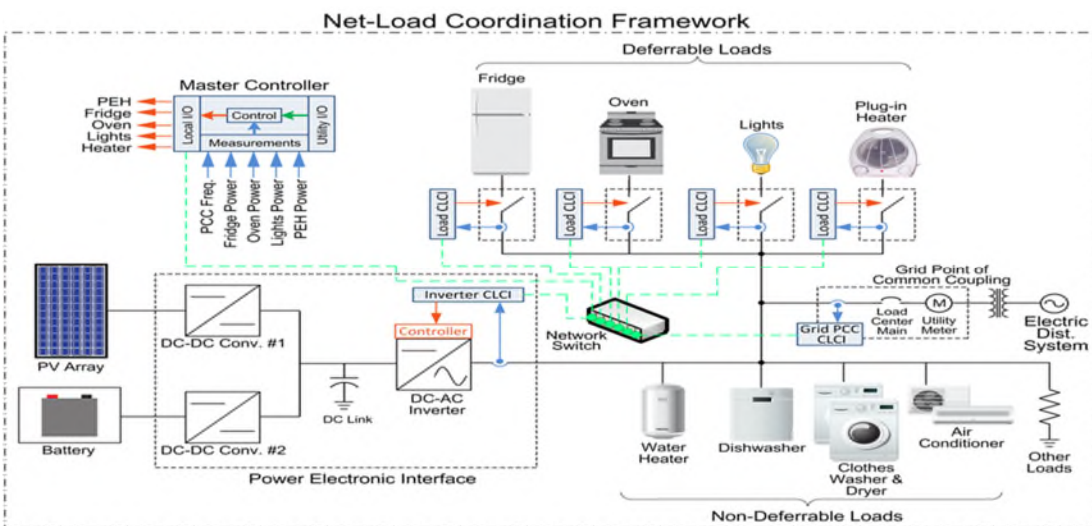


Figure 3-2. Net-Load Coordination Framework Example

Residential-scale power system with net load control. The user determines deferrable and non-deferrable load categories.[92] Notes: CLCI is control layer communication interface; PCC is point of common coupling.

[90] See http://www.obix.org/.
[91] Narang et al., forthcoming.
[92] Lundstrom et al. 2018, Figure 1.

3.2.3. Privacy and cybersecurity

Privacy and cybersecurity are becoming increasingly important as more buildings provide grid services and more end uses are interconnected in those buildings and to the grid.[93] In addition, as assessments rely on the collection of customer data, privacy and cybersecurity also are key considerations for ensuring the safety and security of such assessments and the communications used to collect the data, as well as the underlying data itself. Each communication interconnection is a potential point of privacy intrusion or cyberattack. Thus, privacy and cybersecurity protocols will be needed throughout the chain of systems that enable demand flexibility, from devices, building management systems, and meters to utilities, regional grid operators, aggregators, and third-party evaluators. The increased flow of customer data also may require more attention from regulators to maintain privacy, including a focus on how customers control access by third-party vendors to their utility data.[94]

Cybersecurity is critical for reliability and safety of the grid. While OpenADR2.0 and IEEE 2030.5 are both encrypted, which is fundamental to both cybersecurity and privacy, translators embedded within DERs will be necessary to enable end-to-end encryption. While currently there are no widely accepted common standards for end-to-end DER cybersecurity, this is an area of active development.[95]

The National Institute of Standards (NIST) has developed a framework for evaluating risks in a connected grid: NIST 7628.[96] The standard is being reviewed to determine if it is sufficient given advances in DER functionality. GMLC research found that current coordination on cybersecurity risks is insufficient among stakeholders, including utilities, DER vendors, and service providers. Organizing stakeholders to address this issue is an important area for state and local government involvement.

Ongoing work by national energy laboratories and industry may provide some insight into cybersecurity issues. For example:

- Sandia National Laboratory and SunSpec Alliance are developing best practices for DER cybersecurity and have an ongoing working group to develop best practices.[97]
- Lawrence Livermore National Laboratory is working with Hawaiian Electric to evaluate risks related to DER interconnection in Hawaii.[98]
- Pacific Northwest National Laboratory recently issued a report, *Challenges and Opportunities to Secure Buildings from Cyber Threats*.[99]
- IEEE has begun a revision to its 1547.3 standard that is expected to include best practices for DER cybersecurity.
- The Smart Electric Power Alliance has set up a working group on cybersecurity.[100]

3.2.4. Third-party evaluators

Settlements for demand flexibility services in the future will use standard assessment methodologies and automated processes administered by utilities, regional grid operators, and DER aggregators. In a connected and automated future, third-party evaluators will still play an important role in assessing performance of demand

---

[93] A 2018 survey showed that concern over cybersecurity is the number one barrier to the adoption of Internet-of-Things (IoT) technologies by enterprises. See Bain & Company 2018. Also see U.S. DOE 2017b.

[94] Navigant 2015.

[95] Narang et al., forthcoming. Also see U.S. DOE Office of Cybersecurity, Energy Security, and Emergency Response: https://www.energy.gov/ceser/office-cybersecurity-energy-security-and-emergency-response.

[96] See https://csrc.nist.gov/publications/detail/nistir/7628/rev-1/final. Also see https://www.nist.gov/topics/cybersecurity for NIST work on other cybersecurity topics.

[97] See https://sunspec.org/cybersecurity-work-group/.

[98] Narang et al., forthcoming.

[99] See Reeve at al. 2020

[100] See https://sepapower.org/community/member-committees-and-working-groups/cybersecurity-working-group/.

flexibility strategies and technologies, accuracy of settlement methodologies for dispatchable programs, and impacts of non-dispatchable tariffs. As occurs now, grid operators and regulators may continue to require third-party verification to audit programs and build confidence in the validity of assessments and services they evaluate.

Since demand flexibility often requires real-time or near real-time assessment results, third-party verification efforts may consist of verifying the accuracy of metering, data management procedures, and impact calculation methods/assumptions prior to implementation of events and at regular intervals. This would help provide confidence in the immediately reported results. In addition, after-the-fact auditing of the impacts may be conducted for a final validation.

Use of third parties for these purposes will likely be applicable both in regions with vertically integrated utilities and areas with centrally organized wholesale electricity markets. Work will focus on entities responsible for implementing programs, including utilities, regional grid operators, or DER aggregators. Building operators also may hire third-party evaluators to determine whether demand flexibility investments were worthwhile and to optimize performance.

Confidence in accuracy of assessments is necessary for building owners, operators, and occupants to adopt demand flexibility and utilities and regional grid operators to use it to meet grid services. Third-party evaluators, therefore, should meet qualification standards and be independent of the entities being assessed. Evaluator training and certification may instill further confidence in assessments.[101]

3.3. Assessment Approaches for Modulation

Modulation involves the management of active or reactive power in response to signals from grid operators (automatic management) or measurements of grid parameters (autonomous management). Such management can provide frequency response or frequency regulation reserves through load management or power generation (including battery discharge). Modulation also can provide voltage support by adjusting loads that consume reactive power (e.g., induction motors) or by using inverter controls (e.g., in batteries or PV inverters). Of the four demand flexibility modes, modulation has the least amount of experience in terms of use and assessment. At the same time, there are significant opportunities for modulation to provide grid ancillary services.[102] The few programs that have encouraged modulation have supported projects that installed devices and infrastructure that can support rapid changes in a building's active power consumption in response to an automatic generation control (AGC) signal from a grid operator.

Modulation could be implemented in a variety of ways in the future. How it is implemented and what services it provides will strongly influence assessment requirements. Following are possible future scenarios for demand flexibility modulation services.

- For frequency response and regulation services:
    - A participant in the regulation reserve market receives an AGC signal from the grid operator. The AGC signal reflects the operator's imbalance. The participant might be managing a single building or, more likely, aggregating multiple buildings or sub-metered and controlled devices within multiple buildings. For each building and load managed, the participant predetermines how much flexibility could be provided from changes in load and how much generation could be provided by on-site generators/storage. The building response would be automatic—

[101] For an example of a training and certification program for efficiency impact evaluations, see https://www.aesp.org/general/custom.asp?page=EvaluationCertificateProgram.

[102] "Utility grid operators are continuously monitoring grid frequency and may benefit from the ability to modulate certain electrical loads as a means to maintain the nominal system frequency of 60 Hz in the United States. By modulating a building's electrical load at the second to sub-second level, utilities could impact grid frequency in real time. These frequency adjustments may be necessary when larger electrical loads and generation capacity come on or off line." U.S. GSA 2019, footnote 4. Baker et al. 2016; Lundstrom et al. 2018.

implemented without operator interaction by the building's BAS or individual devices. This would be similar to the shed, shift, and generate modes in terms of the wide range of possible assessment metrics and measurements, but on a shorter timescale, on the order of seconds to minutes, and possibly without the need for baselines.

  - A building may monitor grid frequency at the point of interconnection and respond accordingly by adjusting load or power output from distributed generation/storage systems. This could support a grid operator meeting its frequency response obligation. The building responds autonomously, through controlled devices, without a signal from the grid operator. Assessment metrics and measurements would be similar to those for shed, shift, and generate modes, but on a much faster timescale—sub-seconds to seconds—and may not require baselines.

- For reactive power (voltage support):[103]
  - A participating building continuously monitors voltage levels at the point of interconnection with the grid and responds to voltage excursions by managing reactive power, potentially using smart inverters. Fine-tuning of voltage control settings may be necessary so that controls of neighboring facilities on a feeder do not conflict with each other.
  - A BAS in a participating building detects an issue on the grid, such as a contingency event or abnormal operation.[104] The BAS responds by shutting off certain equipment, such as motors, that can consume significant amounts of reactive power or implementing fast-response reactive power measures (e.g., through smart inverters).

In the future when buildings regularly provide modulation-related grid services under one or more of these scenarios, utilities and regional grid operators will need to determine standardized rules for participation and assessment protocols, as they have for other grid services. Assessment challenges are related to modulation's impact at intervals of sub-seconds to seconds, defining appropriate metrics, and determining whether counterfactual scenarios are required. Following are specific issues and options.

- *Speed*—Time frames and measurement units associated with modulation require sophisticated metering, communication, and interoperability infrastructure at the building level. This may include metering (e.g., utility metering at the building interconnection point or building devices with embedded metering and data transmission capabilities) that measures and reports true power and reactive power demand—depending on the services provided—in very short intervals. If the response time frame is very short, no more than one or two seconds, the building may need to have its own grid monitoring capability, as there would not be time for a grid operator to monitor and signal the building that corrections are required. Information needs to be transmitted for recording, assessment analysis, and confirmation of implemented system or equipment control sequences. Transmittal for assessment likely also will require advanced telemetry for buildings to communicate with aggregators and grid operators in short time frames, emphasizing the need for interoperability and communication protocols as discussed in Section 3.2
- *Metrics for frequency support*—The factors used to assess modulation performance could be simple or complex. For example:

[103] Grid operators currently do not rely on buildings to regulate voltage, although they may charge large customers higher costs for low power factors. See https://www.electronics-tutorials.ws/accircuits/reactive-power.html for a summary discussion of reactive power and power factors.

[104] For example, a fault-induced delayed voltage recovery (FIDVR), an unexpected delay in the recovery of voltage to its nominal value following the normal clearing of a fault. See https://certs.lbl.gov/initiatives/fidvr.

- Whether the building responded to the grid operator's request for modulation services by taking a particular action, such as whether batteries discharged when requested (see Hawaii text box in Section1)
- A performance score that indicates how closely the building's power draw followed the grid signal for a certain load shape (profile) each hour
- The amount of time to full response (e.g., time to 90% of kW set point)[105]
- The amount power consumption, at the building interconnection point or for specific building devices, increased and decreased (+kW, -kW) relative to a specified amount of demand.

CAN EXPERIENCE WITH CONSERVATION VOLTAGE REDUCTION SUPPORT DEVELOPMENT OF MODULATION ASSESSMENT PROTOCOLS?

Conservation Voltage Reduction (CVR) involves optimizing distribution voltage levels during peak periods to achieve peak demand reductions, reducing energy losses in the electric distribution system, or reducing voltage levels for longer periods to achieve electricity conservation (U.S. DOE 2012). While CVR does not involve modulating building loads, it is similar in concept and has some assessment experience. M&V protocols for CVR could serve as starting points for assessing modulation; however, they are works in progress, with issues at least for the efficiency mode with respect to defining baselines (see Northwest Public Power Planning Council, Regional Technical Forum, Automated CVR Protocol No. 1, https://rtf.nwcouncil.org/standard-protocol/automated-cvr-protocol-no-1).

- *Metrics for active or reactive power (voltage) support*—These factors also may be simple or complex. For example:
  - Whether the building responded to the grid operator's request for modulation services by taking a particular action, such as shutting down devices with high reactive power consumption
  - Power factor at the building interconnection point, during an event, is within a predefined power factor range
  - Power factor at the building interconnection point, during an event, relative to a baseline power factor
  - System average voltage indices; for example, whether voltage is within specified bounds (e.g., a violation index).[106]
- *Counterfactuals*—Assessing modulation may or may not require a baseline. For example, the performance metric for controlling how much reactive power a building draws may simply be defined by whether the building's reactive power demand (kVAR) is within a defined range per a tariff or contractual arrangement between a building owner and a servicing utility (or the utility's representative, such as an aggregator). Similarly, for ramping demand up or down, the assessment metric may simply focus on whether load ramped up or down as expected, including within the specified time frame. Alternatively, modulation for ramping may be assessed just as load shedding is typically assessed, using a counterfactual scenario.
- *Assessment boundaries*—The assessment boundary may be critical. While grid services provided by individual devices may indicate that the appropriate boundary is at the device level, interactive effects of other loads, generation, and storage within a building may reduce or amplify the actual impacts at the point of interconnection with the grid. Thus, such interactive effects may instead indicate that the assessment boundary should be at the building level.

[105] For example, see California Independent System Operator Resource Performance Verification Operating Procedure, 5370, Version 7.0, March 11, 2020, 5. http://www.caiso.mobi/Documents/5370.pdf.
[106] See Ding et al. 2016. Section 3.2.

Given the small number of current applications of rapid-response modulation, the above issues and potential metrics are speculative at this time and also are dependent on actual modulation services provided by a building. As the use of modulation is expanded, further research is needed to confirm appropriate metrics, as well as development and implementation of analysis protocols, advanced metering, and interoperable communication systems that can perform with the speed that is fundamental to modulation services. The fast time performance is perhaps the greatest challenge, requiring effective metering, communications, and interoperability, as discussed in the prior subsection, at the building and grid interface, and potentially also within the building itself.

4. Prioritized Actions State and Local Governments Can Take to Advance Demand Flexibility Assessments

State and local governments can play important roles to support reliable and cost-effective performance assessments for demand flexibility (Table 4-1), as well as for energy efficiency. They can encourage assessments, share results, and support adoption of current best practices and advancements. State and local governments also can help overcome assessment challenges identified in this report (Section 3) as more sophisticated forms of demand flexibility are increasingly implemented. Such steps may include improving access to data from existing utility billing meters and considering improvements in metering infrastructure and related standards and protocols.

Table 4-1. Actions State and Local Governments Can Take to Support Demand Flexibility Assessments

Types of Actions	State and Local Government Roles			
	Regulate Jurisdictional Utilities	Operate Demand Flexibility Programs	Establish Codes and Standards	Operational Responsibilities for Public Buildings
Encourage assessments and share results	X	X		X
Adopt current best practices	X	X	X	X
Support advances in assessment practices	X	X		
Improve access to data from existing utility billing meters	X	X		
Consider improvements in metering infrastructure and related standards and protocols (including for utility service connections)	X	X	X	

Among relevant state and local government agencies are those that:

- Regulate utilities (PUCs) or operate demand response and other demand flexibility programs for buildings (e.g., publicly owned utilities), with opportunities to:
  - Establish assessment requirements for programs
  - Directly incent adoption of "demand flexibility-ready" equipment and systems[107] or indirectly incent them through pay-for-performance programs
  - Provide technical support for building owners and managers
  - Require use of specific communication standards to improve interoperability and interconnection standards to facilitate the use of DERs in GEBs

[107] Demand flexibility-ready means equipment and systems that are capable of varying their electricity demand in response to signals from building operators, aggregators, or grid operators through some form of automation, from simple on/off controls to sophisticated control algorithms with feedback.

- Set appliance and equipment standards and building energy codes[108] for new construction and major renovations and, in some jurisdictions, establish benchmarking or retrofit upon sale criteria for privately owned buildings, each with opportunities to advance demand flexibility- ready[109] equipment and building controls (e.g., state energy offices, state and local building codes agencies).[110]
- Have operational responsibilities for public buildings and can implement state-of-the-art assessment practices and lead by example projects, pilot programs, and demonstration testbeds (e.g., administrative services agencies and state and local energy offices).[111]

Stakeholder engagement by utilities, regional grid operators, aggregators, DER providers, and building owners and operators can support these types of state and local government activities. Working collaboratively, they can determine the necessary actions to develop rigorous assessments of demand flexibility at reasonable cost. In addition, critical technical challenges such as cybersecurity and privacy protocols, interoperability standards, and grid operator visibility into distribution systems and buildings are all the subject of ongoing research by DOE,[112] its national laboratories,[113] and industry organizations such as the Electric Power Research Institute (EPRI).[114]

Following is a list of actions state and local governments can consider, in order of priority and ease of implementation, to advance robust performance assessments of demand flexibility in their jurisdictions.

1. Encourage performance assessments, provide technical assistance, and share results

As a first step, state and local governments can lead by example by requiring ongoing performance assessments of demand flexibility for public buildings they own or manage.[115] They also can provide technical assistance to building owners and operators on how to effectively implement and assess demand flexibility. Such technical assistance can be provided by government staff or consultants, as is the case for benchmarking programs.[116]

An important next step is disseminating assessment results and the metrics, analysis methods, and implementation strategies used. State and local governments also can encourage or require performance assessments for buildings participating in dispatchable programs they regulate or operate and share results. In addition, performance assessments are appropriate for pilot programs and full-scale adoption of non-dispatchable time-varying retail rates, to test and refine tariff design and estimate impacts. Conducting assessments, leveraging data, and sharing results can:

- Increase understanding and continuous improvement of demand flexibility performance for the benefit of building owners and occupants and acquisition of grid services at least cost for utility customers
- Inform DER potential studies and electricity system planning based on verified performance
- Secure confidence in assessment practices and demand flexibility as a grid resource.

---

[108] State and local agencies also can adopt national standards, such as ENERGY STAR®, for appliances, equipment, buildings, and smart home energy management systems, as well as for benchmarking building energy and demand.
[109] George 2010.
[110] U.S. DOE 2019b.
[111] See American Council for an Energy-Efficient Economy (ACEEE), State Government Lead by Example website: https://www.aceee.org/state-policy-toolkit. Also see U.S. Environmental Protection Agency's website on Energy Resources for State, Local, and Tribal Governments: https://www.epa.gov/statelocalenergy.
[112] Neukomm et al. 2019.
[113] https://gmlc.doe.gov/.
[114] See https://www.epri.com/#/portfolio/2020/research_areas/0/000031?lang=en-US.
[115] Thresholds for building size or peak demand may be appropriate for such a requirement.
[116] Mims et al. 2017, chapters 3 and 6.

2. Adopt current best practices for demand flexibility assessments

Initially, state and local governments can incorporate current best assessment practices for performance metrics, analysis methods, baselines, and implementation for their own programs or pilot projects.[117] For example, technical reference manuals[118] for energy efficiency programs or demand response M&V manuals[119] could be expanded to incorporate information on demand flexibility from GEBs. Ultimately, jurisdictions should consider adopting guidance for performance assessments for both dispatchable demand flexibility programs and non-dispatchable time-varying tariffs.

State and local governments also can support standardized protocols for demand flexibility assessments that can reduce costs and increase consistency and credibility of assessment findings. This includes standardized data-sharing protocols, discussed in Section 3 and below. Related, designing programs and policies that leverage metering and regular reporting of GEB performance as part of implementation can reduce costs for performance assessments and provide results quicker. See Sections 1 and 2 and Appendix C for examples of best practices examples and resources.

Further, to help ensure consistency in the valuation of DERs and their demand flexibility attributes using a common benefit-cost analysis framework, state and local governments can refer to the National Standard Practice Manuals (NESP 2017 and NESP forthcoming) to guide their cost-effectiveness analyses.

For future advances in demand flexibility, best practices may be developed through state and local government initiatives and standardization efforts by utilities, regional grid operators, standards organizations, and federal initiatives. Enabling participation by third-party providers in pilots and programs also can facilitate innovative approaches.[120]

3. Support advances in assessment practices

As continuous demand flexibility becomes more common, assessment practices will need to improve in several areas:

- Performance metrics[121]
    - Building service impact metrics (e.g., affordability, comfort, and indoor air quality), considering perspectives of building owners, operators, and occupants for both publicly and privately owned buildings
    - Quantity and quality performance metrics to better address the modulation flexibility mode, integrated DERs, and multiple flexibility modes to provide a broader range of grid services to utilities and centrally organized wholesale electricity markets
- Assessment strategies, including baselines as warranted,[122] for determining the quantity, quality, and value of grid services provided by:
    - Combined load shed and shift
    - Buildings providing grid services for more than one dispatchable program
    - Modulation

[117] See U.S. GSA 2019 for an example of a program to assess buildings and technologies to cost-effectively provide load flexibility.
[118] SEE Action 2017.
[119] Goldberg and Agnew 2013.
[120] For example, see https://www.energy.gov/eere/buildings/grid-interactive-efficient-buildings.
[121] Liu 2019.
[122] In particular, assessment strategies for counterfactual approaches for use with continuous or near-continuous demand flexibility (see Section 3).

- Integrated DERs providing multiple flexibility modes

- Assessment implementation
  - Development and application of cybersecurity and privacy protocols and interoperability standards
  - Criteria for qualification of evaluators who conduct and report assessment results, including independence and minimizing conflicts of interest.

State energy offices in particular can support these advances through researching and sharing assessment practices; establishing pilot programs for public buildings that demonstrate, though documented assessments, demand flexibility capabilities as well as cybersecurity and privacy protocols; and encouraging or requiring application of the most current interoperability standards, including optional elements of the standards that support demand flexibility operation and assessments. State agencies also can use assessment results to indicate the potential and role of demand flexibility in statewide energy plans and building energy code and appliance standard updates.

4. Improve access to data from existing and future utility billing meters

Even where utilities have deployed AMI, customers may not have ready access to hourly (or more granular) data that these systems may record. In addition, third-party DER program implementers and other service providers may not have ready access to such data, even where the customer provides consent. State and local governments can adopt data access provisions, protocols, or standards to address such issues where present.

5. Consider improvements in metering infrastructure and related standards and protocols

State and local governments interested in modernizing utility distribution systems, including metering infrastructure, can first identify motivating objectives.[123] One potential objective is using DERs to help meet generation and T&D needs. That may lead to PUC approval of a regulated utility's proposal (or guidance from the PUC to the utility, or from management or the board for publicly owned utilities and rural electric coops) to improve capabilities for metering infrastructure—specifically, implementation of AMI that meets current IEEE standards.[124] Such a system could have capabilities to continuously measure and communicate buildings' demand (kW) and energy (kWh) in real time, improving the quality of performance assessments. In addition, for some autonomous modulation services, systems could measure reactive power (kVAR), grid frequency (Hz), and voltage (V).

The jurisdiction's objectives also will shape investment priorities for customer access to demand flexibility performance information as well as distribution system monitoring.[125] Such information enables higher penetrations of DERs and demand flexibility and supports assessments by providing data on electrical loads for individual and aggregated buildings and distribution feeders for those buildings.

With respect to investments in sub-metering and BAS by building owners, state and local governments can support such improvements in a variety of ways:

[123] "Objectives relate to either defining customer experience or system characteristics that require either changing or improving existing capabilities or adding new capabilities. In this context, an objective is a goal/outcome with an associated timing or performance metric." Objectives and corresponding capabilities typically fall within the categories of reliability/resilience, safety and operational efficiency, and DER integration and/or DER utilization. U.S. DOE 2019c, 28.
[124] For related recommendations, see the Technology, Metering, Communications, and Data Systems section in Shenot et al. 2019.
[125] U.S DOE 2019a.

- Leading by example by taking such actions in their own buildings and sharing results
- Encouraging sub-metering capabilities and BAS enhancements through rebates and other incentives
- Encouraging or requiring sub-metering and demand flexibility-ready equipment for certain end uses in large buildings through codes and standards.[126]

All of these metering advancements provide greater access to data needed for quality performance assessments. At the same time, these advancements raise issues related to communication interoperability, cybersecurity, and protection of sensitive building, personal, and organizational information. State and local governments can support adoption and implementation of standards and protocols for interoperability, cybersecurity, and privacy.

6. Improve coordination across programs and markets and simplify implementation of demand flexibility

Programs and markets operated by different entities within a jurisdiction—utilities, RTOs/ISOs, and state and local governments—often lack coordination. In addition to reduced participation and performance, this can lead to double-counting and conflicting rules, roles, and responsibilities. Governor's offices, state energy offices, and city councils can initiate efforts to improve coordination and communication across entities and the various programs and markets that allow participation by DERs, in concert with integrated demand-side management and GEB concepts. Such facilitation support can help align priorities and objectives across stakeholders, leading to broadly supported actions.[127]

Buildings providing demand flexibility in increasingly sophisticated ways can increase operational complexity, creating potential barriers for program implementation by utilities, regional grid operators, and aggregators and participation by consumers. Diverse standards, policies, rates, and wholesale market design can add to this complexity. Increased coordination of programs and markets can encourage adoption of simplified implementation approaches.

State and local governments also may wish to perform market evaluations for energy efficiency and demand flexibility services in their jurisdictions, as part of market transformation activities.

[126] For example, the California building energy code requires that certain buildings are demand response-ready (i.e., capable of responding to a demand response signal). See https://www.energy.ca.gov/sites/default/files/2020-05/Appendix_D_DemandResponsiveControls.pdf.

[127] For additional information on state and local government actions that can advance demand flexibility, see Table 4 and Appendix C in SEE Action Network 2020a.

5. Conclusions

Performance assessments determine the quantity and quality of grid services provided. For GEBs, assessment information can be used to improve demand flexibility performance, assess its impact and cost-effectiveness, and support its consideration in electricity system planning. State and local governments can take actions in partnership with stakeholders to advance practices for demand flexibility performance assessments to support the jurisdiction's own energy objectives related to electricity system reliability and resilience, energy affordability, and integration of new electric loads and resources.

To a large degree, current best practices are sufficient for assessing basic grid services currently provided through demand flexibility. However, advances in assessment practices will be required in a future with GEBs providing continuous demand flexibility by integrating multiple DERs and demand-side strategies (efficiency, load shed, load shift, modulate, and generate). Example practices include:

- Using new baseline constructs—potentially including no baselines for certain building operating modes or configurations
- Deploying more advanced metering and analytics
- Further developing cybersecurity standards
- Improving communication standards for increased interoperability
- Establishing performance metrics and assessment procedures for load modulation to address its fast response nature and the grid services it can provide
- Simplifying overall approaches to demand flexibility implementation and performance assessments

References

AEIC (Association of Edison Illuminating Companies), Pacific Gas & Electric Company, Southern California Edison Company, San Diego Gas & Electric Company. 2018. *Enabling Smart Inverters for Distribution Grid Services*. Joint IOU White Paper Series. October. https://www.pge.com/pge_global/common/pdfs/about-pge/environment/what-we-are-doing/electric-program-investment-charge/Joint-IOU-SI-White-Paper.pdf.

ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). 2014. *Guideline 14-2014—Measurement of Energy, Demand, and Water Savings*.

Baker, Kyri, Xin Jin, Deepthi Vaidhynathan, et al. 2016. *Short Paper: Frequency Regulation Services from Connected Residential Devices*. National Renewable Energy Laboratory. November. https://www.nrel.gov/docs/fy17osti/66586.pdf.

Bain & Company. June 13, 2018. "Cybersecurity Is the Key to Unlocking Demand in the Internet of Things." https://www.bain.com/insights/cybersecurity-is-the-key-to-unlocking-demand-in-the-internet-of-things/.

Balaji, Bharathan. 2018. "Brick: Metadata Schema for Portable Smart Building Applications." *Applied Energy* 226: 1273–1292. September. DOI: 10.1016/j.apenergy.2018.02.091.

Best, Carmen, Megan Fisher, and Mark Wyman. 2019. "Policy Pathways to Meter-Based Pay-for-Performance." International Energy Program Evaluation Conference. https://www.iepec.org/2019_proceedings/#/paper/event-data/044-pdf.

Bushby, Steve T. 2011. "Information Model Standard for Integrating Facilities with Smart Grid." *ASHRAE Journal* November.

California ISO. 2017. *Baseline Accuracy Work Group Proposal*. June. https://www.caiso.com/Documents/2017BaselineAccuracyWorkGroupFinalProposalNexant.pdf.

California ISO. 2018*. Demand Response User Guide*. Version 4.5. California ISO. http://www.caiso.com/Documents/DemandResponseUserGuide.pdf.

Cappers, Peter, Andrew Mills, Charles Goldman, Ryan Wiser, and Joseph H. Eto. 2011. *Mass Market Demand Response and Variable Generation Integration Issues: A Scoping Study*. Lawrence Berkeley National Laboratory. https://eetd.lbl.gov/sites/all/files/publications/lbnl-5063e.pdf.

ConEdison. 2020. "Energy Efficiency and Demand Management Procedure – General Calculating Customer Baseline Load." Specification EE-DR-01. Revision 03. Effective date January 1 2020. https://www.coned.com/-/media/files/coned/documents/save-energy-money/rebates-incentives-tax-credits/smart-usage-rewards/customer-baseline-load-procedure.pdf?la=en.

Dayem, Katherine. 2018. *Standardized Communications for Demand Response*. National Rural Electric Cooperative Association (NRECA). June. https://www.cooperative.com/programs-services/bts/documents/reports/standardized-communications-for-demand-response-report-june-2018.pdf.

Demand Resources Working Group. 2019. "Assessing Energy Efficiency Resource Performance in All Hours." NEPOOL Markets Committee. https://www.iso-ne.com/static-assets/documents/2019/07/reportofthedrwgonassessingeerperformanceinallhours_final2.pdf.

Ding, Fei, Adarsh Nagarajan, Sudipta Chakraborty, and Murali Baggu, National Renewable Energy Laboratory, and Andu Nguyen, Sarah Walinga, Michael McCarty, and Frances Bell, SolarCity. 2016. *Photovoltaic Impact Assessment of Smart Inverter Volt-VAR Control on Distribution System Conservation Voltage Reduction and Power Quality.* NREL/TP-5D00-67296. December. https://www.nrel.gov/docs/fy17osti/67296.pdf.

ERCOT (Electric Reliability Council of Texas). 2019. *Demand Response Baseline Methodologies Version 19.0*. September. http://www.ercot.com/services/programs/load/eils/documents.

Faruqui, Ahmad, Sanem Sergici, and Lamine Akaba. 2014. "The Impact of Dynamic Pricing on Residential and Small Commercial and Industrial Usage: New Experimental Evidence from Connecticut" *The Energy Journal* 35 (1).

Faruqui, Ahmad, Neil Lessem, and Sanem Sergici. 2017. "Dynamic Pricing Works in a Hot, Humid Climate." *Public Utilities Fortnightly.*

FERC (Federal Energy Regulatory Commission). 2014. ISO New England Inc. and New England Power Pool, Order on Tariff Filing and Instituting Section 206 Proceeding, 147 FERC ¶ 61,172. Paragraph 89, https://www.iso-ne.com/regulatory/ferc/orders/2014/may/er14-1050-000_5-30-14_pay_for_performance_order.pdf.

FERC. 2018. *Assessment of Demand Response and Advanced Metering.* Staff Report. November. https://www.ferc.gov/reports-demand-response-and-advanced-metering.

Fernandez, N, S. Katipamula, W. Wang, Y. Xie, M. Zahao, and C. Corbin. 2017. *Impacts of Commercial Building Controls on Energy Savings and Peak Load Reduction.* Pacific Northwest National Laboratory. PNNL-25985. May. https://buildingretuning.pnnl.gov/publications/PNNL-25985.pdf.

Franconi, Ellen, Matt Gee, Miriam Goldberg, Jessica Granderson, Tim Guiterman, Michael Li, and Brian Arthur Smith. 2017. "The Status and Promise of Advanced M&V: An Overview of 'M&V 2.0' Methods, Tools, and Applications." LBNL-1007125, 1350974. https://doi.org/10.2172/1350974.

George, Karen. 2010. "Functional Requirements for Demand Response Ready Appliances." Electric Power Research Institute. EnergyStar Products Partner Meeting. October.

Goldberg, Miriam L., and G. Kennedy Agnew. 2013. *Measurement and Verification for Demand Response.* National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group. February. https://emp.lbl.gov/publications/measurement-and-verification-demand.

Granderson, Jessica, and Samuel Fernandes. 2017. "State of Advanced Measurement and Verification Technology and Industry Application." *The Electricity Journal* October, 8–16.

Green Mountain Power. 2019. GMP Bring Your Own Device 'BYOD' Access & Service Agreement. March. https://greenmountainpower.com/wp-content/uploads/2019/03/BYOD-Terms-and-Conditions-3-11-19.pdf.

IEEE 2011. *Standard 2030-2011 - IEEE Guide for Smart Grid Interoperability of Energy Technology and Information Technology Operation with the Electric Power System (EPS), End-Use Applications, and Loads*. https://standards.ieee.org/standard/2030-2011.html.

IEEE 2018. Standard 1547-2018 - IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces. https://standards.ieee.org/standard/1547-2018.html.

ISO New England. 2018. *Measurement and Verification of On-Peak Demand Resources and Seasonal Peak Demand Resources.* ISO New England. https://www.iso-ne.com/static-assets/documents/2018/10/manual_mvdr_measurement_and_verification_of_onpeak_and_seasonal_peak_demand_resources_rev07_20181004.pdf.

ISO New England. 2019. Market Rule 1, Section III: Standard Market Design. October. https://www.iso-ne.com/static-assets/documents/2014/12/mr1_sec_1_12.pdf.

Li, Michael, Hossein Haeri, and Arlis Reynolds. 2018. *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. National Renewable Energy Laboratory. August. https://www.nrel.gov/docs/fy18osti/70472.pdf.

Liu, Jingjing. 2019. "Framework & Method to Define Flexible Loads in Buildings - to Integrate as a Dynamic & Predictable Grid Resource." Presentation to the Technical Advisory Group meeting, June 26, Lawrence Berkeley National Laboratory.

Lundstrom, Blake, Sourav Patel, Sandeep Attree, and Murti V. Salapaka. 2018. *Fast Primary Frequency Response Using Coordinated DER and Flexible Loads: Framework and Residential-Scale Demonstration.* National Renewable Energy Laboratory. July. https://www.nrel.gov/docs/fy18osti/70505.pdf.

Luo, Xuan, Tianzhen Hong, Yixing Chen, and Mary Ann Piette. 2017. *Electric Load Shape Benchmarking for Small- and Medium-Sized Commercial Buildings*. Lawrence Berkeley National Laboratory. March. http://eta-publications.lbl.gov/sites/default/files/t_hong_-_electric_load_shape_benchmarking_for_small-_and_medium-sized_commercial_buildings.pdf.

Mass Save. 2019. *Program Materials for Connected Solutions for Small Scale Batteries*. Version 17, https://www.masssave.com/-/media/Files/PDFs/Save/Residential/connectedsolution-batteries/Program-Materials-for-ConnectedSolutions-for-Small-Scale-Batteries.pdf?la=en&hash=90F93C9BC4BFBD8D6160CCAA5E594A64B6C8B29B.

MISO (Midcontinent System Operator). 2019. "Business Practices Manual Demand Response." Business Practices Manual BPM-025-r4. https://www.misoenergy.org/legal/business-practice-manuals/.

Mims, Natalie, Steven R. Schiller, Elizabeth Stuart, Lisa Schwartz, Chris Kramer, and Richard Faesy. 2017. *Evaluation of U.S. Building Energy Benchmarking and Transparency Programs: Attributes, Impacts, and Best Practices*. Lawrence Berkeley National Laboratory. April. https://eta.lbl.gov/publications/evaluation-us-building-energy.

Narang, David, Peter Schwartz, Steve Widergren, Sigifredo Gonzalez, S. M. S. Alam, Theodore Bohn, Yaosuo Xue, et al. Forthcoming. *GMLC Gap Analysis for Distributed Energy Resource Interconnection and Interoperability Standards and Test Procedures.* Prepared for U.S. DOE Grid Modernization Laboratory Consortium.

NESP (National Efficiency Screening Project). 2017. *National Standard Practice Manual for Assessing Cost-Effectiveness of Energy Efficiency Resources*. https://nationalefficiencyscreening.org/wp-content/uploads/2017/05/NSPM_May-2017_final.pdf.

NESP. 2020. Database of State Efficiency Screening Practices (DSESP). https://nationalefficiencyscreening.org/state-database-dsesp/.

National Grid and the Energy Efficiency Advisory Council. 2019. *Evaluation of 2018 Demand Response Demonstration: C&I Connected Solutions*. February. http://ma-eeac.org/wordpress/wp-content/uploads/National-Grid-Connected-Solutions-Final-Report-02232018.pdf.

Navigant Consulting. 2015. *Value of Customer Data Access: Market Trends, Challenges, and Opportunities*. Prepared for National Association of Regulatory Utility Commissioners. April. https://pubs.naruc.org/pub.cfm?id=536E2C7B-2354-D714-51CE-F035BA50FAA1.

Navigant. 2018. *RES 1 Baseline Load Shape Study*. Prepared for The Electric and Gas Program Administrators of Massachusetts. http://ma-eeac.org/wordpress/wp-content/uploads/RES-1-FINAL-Comprehensive-Report-2018-07-27.pdf.

Neukomm, Monica, Valerie Nubbe, and Robert Fares. 2019. *Grid-Interactive Efficient Buildings: Overview*. U.S. DOE. December. https://www1.eere.energy.gov/buildings/pdfs/75470.pdf.

Newman, Michael H. 1997. "BACnet: Answers to Frequently Asked Questions." http://www.bacnet.org/Bibliography/HPAC-3-97.pdf.

Nordham, Bruce, Margarita Kloss, Bijit Kundu. 2019. *Energy Reporting: Device Demonstration, Communication Protocols, and Codes and Standards.* California Energy Commission. April. https://escholarship.org/uc/item/61j0f2vg.

OpenADR 2.0 Profile Specification B Profile. 2013. OpenADR Alliance.

PJM. 2019. PJM Manual 11: Energy & Ancillary Services Market Operations. Revision 107, September 26. https://www.pjm.com/directory/manuals/m11/index.html#Sections/Introduction.html.

Prairie, Dan, John Petze, Marc Petock, and Mathan Kumar Gopal Samy. 2016. "*Project-Haystack*." Continental Automated Buildings Association (CABA). https://project-haystack.org/file/22/CABA-White-Paper-on-Project-Haystack.pdf.

Reeve, Hayden, Sri Nikhil, Gupta Gourisetti, Penny McKenzie, Michael Mylrea, Paul Ehrlich, Glenn Fink, and Draguna Vrabie. 2020. *Challenges and Opportunities to Secure Buildings from Cyber Threats*, Pacific Northwest National Laboratory, PNNL-29813, March. https://www.energy.gov/sites/prod/files/2020/05/f74/bto-pnnl-29813-securing-buildings-cyber-threats-051420.pdf.

Schwartz, L. C., Max Wei, William R. Morrow III, Jeff Deason, Steven R. Schiller, Greg Leventis, Sarah Josephine Smith, Woei Ling Leow, Todd Levin, Steven Plotkin, Yan Zhou, and Joseph Teng. 2017. *Electricity End Uses, Energy Efficiency, and Distributed Energy Resources.* January. https://emp.lbl.gov/publications/electricity-end-uses-energy.

SEE Action Network. 2012. *Energy Efficiency Program Impact Evaluation Guide.* Prepared by Steve R. Schiller, Lawrence Berkeley National Laboratory. https://www4.eere.energy.gov/seeaction/system/files/documents/emv_ee_program_impact_guide_0.pdf.

SEE Action Network. 2017. *Guidance on Establishing and Maintaining Technical Reference Manuals for Energy Efficiency Measures*. Prepared by Steve R. Schiller, Greg Leventis, Tom Eckman, and Sean Murphy, Lawrence Berkeley National Laboratory. https://www4.eere.energy.gov/seeaction/system/files/documents/TRM%20Guide_Final_6.21.17.pdf.

SEE Action Network. 2020a. *Grid-Interactive Efficient Buildings: An Introduction for State and Local Governments.* Prepared by: Lisa Schwartz and Greg Leventis, Lawrence Berkeley National Laboratory. https://www.energy.gov/sites/prod/files/2020/04/f74/bto-see-action-GEBs-intro-20200415.pdf.

SEE Action Network 2020b. Determining Utility System Value of Demand Flexibility from Grid-Interactive Efficient Buildings. Prepared by: Tom Eckman, Lisa Schwartz, and Greg Leventis, Lawrence Berkeley National Laboratory. https://www.energy.gov/sites/prod/files/2020/04/f74/bto-see-action-GEBs-valuation-20200410.pdf.

Seymour, Joseph. 2012. *The Seven Types of Power Problems*. Schneider Electric. White paper 18. https://download.schneider-electric.com/files?p_Doc_Ref=SPD_VAVR-5WKLPK_EN.

Shenot, J., C. Linvill, M. Dupuy, and D. Brutkoski. 2019. *Capturing More Value from Combinations of PV and Other Distributed Energy Resources*. Regulatory Assistance Project. August. https://www.raponline.org/wp-content/uploads/2019/08/rap_shenot_linvill_dupuy_combinations_pv_other_ders_2019_august.pdf.

Smart Electric Power Alliance and Navigant Consulting. 2017. *2017 Utility Demand Response Market Snapshot*. October. https://sepapower.org/resource/2017-utility-demand-response-market.

Todd, Annika, Peter Cappers, C. Anna Spurlock, and Ling Jin. 2019. "Spillover as a Cause of Bias in Baseline Evaluation Methods for Demand Response Programs." Lawrence Berkeley National Laboratory.

Trabish, Herman. 2019. "Northeastern Utilities Aim to 'Crush and Flatten' System Peaks as DERs Boost Grid Efficiency." *Utility Dive*. September 23. https://www.utilitydive.com/news/northeastern-utilities-aim-to-crush-and-flatten-system-peaks-as-ders-boos/562944/.

U.S. Department of Energy (DOE). 2012. *Application of Automated Controls for Voltage and Reactive Power Management—Initial Results*. Smart Grid Investment Grant Program. December. https://www.energy.gov/sites/prod/files/Voltage%26ReactivePower_Dec2012Final.pdf.

U.S. DOE. 2015. M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0. U.S. Department of Energy Federal Energy Management Program. https://www.energy.gov/sites/prod/files/2016/01/f28/mv_guide_4_0.pdf.

U.S. DOE. 2016. *Best Practices for Providing Whole-Building Energy Data: A Guide for Utilities Energy Data Accelerator*. Better Buildings Program. January. https://betterbuildingssolutioncenter.energy.gov/sites/default/files/attachments/Best%20Practices%20for%20Providing%20Whole-Building%20Energy%20Data%20-%20Guide%20for%20Utilities.pdf.

U.S. DOE. 2017a. *Transforming the Nation's Electricity System: The Second Installment of the QER. Quadrennial Energy Review.* Department of Energy. https://www.energy.gov/sites/prod/files/2017/02/f34/Quadrennial%20Energy%20Review--Second%20Installment%20%28Full%20Report%29.pdf.

U.S. DOE. 2017b. Department of Electricity and Energy Reliability. 2017. *Cybersecure Interconnection of Distributed Energy Resources.* https://www.energy.gov/sites/prod/files/2017/09/f36/LLNL_CyberInterconnection_FactSheet.pdf.

U.S. DOE. 2019a. Grid-interactive Efficient Buildings Technical Report Series, *Whole Building Controls, Sensors, Modeling, and Analytics,* prepared by U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy, Building Technologies Office, and the National Renewable Energy Laboratory. https://www1.eere.energy.gov/buildings/pdfs/75478.pdf.

U.S. DOE. 2019b. Benchmarking and Transparency: Resources for State and Local Leaders. January. https://www.imt.org/wp-content/uploads/2019/02/Benchmarking_Transparency_Resource_PDF_Final_2.14.pdf.

U.S. DOE. 2019c. *Modern Distribution Grid, Guidebook, Volume IV, Draft v5,* sponsored by U.S. Department of Energy's Office of Electricity Delivery and Energy Reliability.

U.S. Energy Information Administration (EIA). 2019. *Annual Energy Outlook 2019*. "Reference Case Projections Tables." https://www.eia.gov/outlooks/aeo/tables_ref.php.

U.S. General Services Administration (GSA). 2019. *Request for Information: 47PA0120N0001*. GSA Proving Ground Program. October. https://www.gsa.gov/governmentwide-initiatives/sustainability/emerging-building-technologies/request-for-information.

Violette, Daniel M., and Pamela Rathbun. 2014. *Chapter 23: Estimating Net Savings: Common Practices, The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures*. National Renewable Energy Laboratory. September. https://www.energy.gov/sites/prod/files/2015/02/f19/UMPChapter23-estimating-net-savings_0.pdf.

Webster, Lia. 2020. *IPMVP's Snapshot on Advanced Measurement & Verification. Efficiency Valuation Organization.* January. http://evo-world.org/images/corporate_documents/NRE-NRA_White_Paper_Final_2701.pdf.

Appendix A. Summary of Interviews with Experts

Berkeley Lab conducted 13 interviews with regional grid operators, utilities, evaluators, demand service providers, and other industry experts as part of its research for this report.

Type of Organization	Organization	Interviewees
Regional grid operator	Independent System Operator (ISO) – New England	Henry Yoshimura
Regional grid operator	Midwest Independent System Operator	Mia Adams, Kristin Swenson, Matthew Tackett, Robert Merring
Regional grid operator	California Independent System Operator	John Goodin
Utility (publicly owned utility)	Sacramento Municipal Utility District	Craig Sherman, Josh Raisin
Utility (investor-owned utility)	Eversource Energy	Michael Goldman and Christopher Chan
Utility research organization	Electric Power Research Institute	Ram Narayanamurthy
Software as a Service Provider (M&V services)	ReCurve Analytics, Inc.	Carmen Best
Program Implementer/Administrator (public)	New York State Energy Research and Development Authority	Alicia Noriega, Michael Reed, and Thomas Yeh
Aggregator/demand response program implementer	OhmConnect, Inc.	Brian Kooiman
Aggregator/demand response program implementer	Enel X N. A.	Thomas Rokholt
Storage/DG company	STEM, Inc.	Ted Ko
Nongovernmental organization	GridWorks	Matthew Tisdale
Consultant and Trade Association Representative	Navigant/California Energy and Demand Management Council	Greg Wikler

We structured the interviews around seven questions that address market and methodological considerations for demand flexibility assessments. Responses tended to focus on the shed mode of demand flexibility, though some experts also addressed the shift, modulate, and generate modes, as well as approaches for assessing efficiency. Following is a summary of responses.

1. *Will demand flexibility assessments differ by market (e.g., vertically integrated utility markets, RTO/ISO markets, commercial versus residential buildings)?*

Interviewees consistently affirmed that demand flexibility assessments will not differ by market structure (e.g., vertically integrated utilities, centrally organized wholesale market, or mixed markets) or customer segment (e.g., residential or commercial) providing the grid service. Instead, the type and objective of the demand flexibility service provided largely determines the nature of the assessment. For example, assessments would be different for energy, capacity, and frequency services.

Responses identified that providing a grid service requires sharing certain information, such as connected load, at certain timescales. One interviewee offered the example of buildings providing frequency regulation services. As with generators, grid operators require no more than a 2- to 4-second response.

The importance of operational objectives also was a theme in our interviews. One interviewee stated that utilities and regional grid operators may have different goals, which could affect their approach to assessments. Utilities in restructured states, for example, may only want to assess the impact of demand flexibility on distribution system congestion. RTOs, on the other hand, are focused on load shapes at the transmission level. Demand flexibility in buildings may provide different services for these two types of entities, requiring different types of assessments.

Responses, however, recognized that the *capability* to provide grid services depends on whether the building is residential or commercial. Interviewees stated that commercial customers are more capable of providing grid services for the following reasons:

- Interval metering (and capability to transmit interval data) may be a technical requirement for performing assessments, and a higher portion of commercial buildings have this type of metering than residential buildings.
- It costs more to acquire residential customers as participants in programs, in terms of the amount of load that can be managed.
- Commercial customers are more likely to have devices capable of providing grid services.

Interviewees also identified that the boundary of an assessment may be different for residential assessments; the small scale of electrical capacity per residential building may require that assessments occur for an aggregation of buildings.

2. How will current M&V practices need to change for the future assessment of demand flexibility?

Interviewees indicated that, while there is always room for improvement, current M&V practices are successfully being used to assess existing demand response programs. For example, for load shedding today, they said current M&V practices are mostly sufficient.

Discussions focused on how assessments might be done, including changes from current practices, in a future with a far greater amount and higher frequency of demand flexibility. Different interviewees focused on different aspects of what might need to change in such a future. Responses identified the following areas where assessment practices may need further development:

- More granular metering data, both in terms of frequency (e.g., interval metering) and boundary (e.g., building or device). One person commented that assessments of modulation in particular would require new meters (as compared to current standard metering) due to the short time frame (i.e., seconds) of service.
- Metrics of demand flexibility given the range of different objectives associated with providing demand flexibility
- Development and deployment of device interoperability
- New baseline approaches given the potential for few non-event days
- Development of assessment methodologies for modulation
- Updating existing M&V methods for shift, perhaps based on M&V methods for shed.

Interviewees frequently mentioned the importance of efforts that would produce more granular and end use-specific metering data. Many identified advanced metering infrastructure (or some other way to get more data) as an enabling condition for future demand flexibility assessments, in particular for shed. A few individuals identified the importance for future assessments of data beyond electricity metering and weather; specifically:

- Occupancy data could be an input for forecasts
- Occupant comfort measurements could provide feedback for building operations
- Power quality (e.g., reactive power) could enable new grid services.

Generally, interviewees stated that the building level—the border between end users and grid operators—is the primary boundary for assessments. However, some interviewees raised the possibility of assessments moving toward devices. Interviewees identified sub-metering as an important way to assess batteries, electric vehicles, and distributed photovoltaic systems. Overall, interviewees conveyed some interest in the shift toward more granular and device-specific assessments, although some were skeptical about the likelihood of disaggregation technologies enabling device-level forecasts. One interviewee noted that well-documented and existing open-source methodologies could increase confidence in, and efficiency of, assessments. Interviewees generally predicted the continued use of third-party evaluators to validate performance, carry out settlement methodologies and perform audits.

Interoperability and communications protocols came up often in our discussions. Interviewees see device interoperability and standardization of communication between grid-interactive efficient buildings and utilities, regional grid operators, and demand service providers as a critical need for accurate and efficient assessments.

Some interviewees identified the need for new assessment metrics. *Quantity* metrics such as kilowatt reduction are important for shed. For frequency regulation services, however, interviewees identified *quality* metrics such as ramp time as important to future demand flexibility assessments.

3. What metrics will be used to assess demand flexibility?

We found general agreement among experts interviewed that the grid services being provided determine the appropriate metrics. One interviewee stated that metrics for a demand flexibility service should convey the information necessary to: (1) demonstrate to the grid operator or market operator that the grid service occurred and (2) serve as the basis for a financial transaction.

Our interviews illustrated how metrics are evolving with grid services. Interviewees recognize that metrics existed for measuring demand response and energy efficiency, namely demand and energy, but see that buildings providing more varied demand flexibility services will require more sophisticated metrics. Still, interview responses often referenced how power draw before and after a signal is a basic performance metric for demand flexibility.

At the same time, many experts said knowing the *quantity* of demand flexibility is not enough, and that metrics for the *quality* of the service also are necessary. The quality metrics they raised often included a time component, such as duration of a demand reduction. A grid operator, for example, may need to know that a building can maintain a level of performance, just as it needs to know a generator can stay online. One interviewee identified that some grid services also would require metrics that capture a resource's ability to respond to signals and bring its demand flexibility capabilities on line. Such metrics would include start-up time and ramp rates.

An additional dimension of the quality of demand flexibility is variation in its quantity. Grid operators need to balance load and supply. Knowing how much a demand reduction could vary over time helps to integrate demand flexibility resources into electricity system planning. One interviewee framed demand flexibility as a DER akin to solar and suggested the use of a metric typically applied to generators: loss of load probability. Other interviewees suggested a more general capacity metric for predicted demand flexibility. For program implementation, one interviewee suggested a metric to track performance of demand response programs: the share of customers subscribed by an aggregator that actively reduces demand.

Some experts discussed the need for metrics for battery performance, such as metered output. To provide this metric, batteries would be metered separately to account for discharges to the grid. One interviewee suggested that occupant metrics such as thermal comfort would be important for building operators.

4. Will demand flexibility assessments cover performance only at the whole-building level? Or will assessments need to cover the performance of each DER, end use, or piece of equipment in the building?

Many of the experts we interviewed expect that most demand flexibility assessments will occur at the building-level, largely because a building's electricity meter is where the data necessary for assessments will be available. Both assessments and the provision of a grid service require transmission of that information. Because the electricity meter already can provide these functions, device-level data may not be necessary for assessments. Interviewees, however, identified sub-metering for PV systems, electric vehicles, and batteries as important exceptions to the building-level assessment boundary, due to their discharges of electricity. If a building meter records the *net* load of a building with a PV system, electric vehicle, or battery, it may mask demand flexibility and make it difficult to assess demand changes. Separate metering for these technologies enables identification of multiple grid services provided by an individual building.

Some interviewees advocated for device-level assessments; for example, identifying its importance to building operators, who may need to fine-tune systems in order to provide grid services. Such device-level data would include more than just energy and demand. Interviewees also pointed out the importance of occupancy measurements throughout a building. In addition, one interviewee said device-level assessments would be necessary for third-party verification and performance validation.

Also raised were barriers to device-level data. First, lack of interoperable devices and communications standards is a barrier to access these data. Second, the inherent variability of device-level data may make its analysis more difficult. Some interviewees raised the possibility of analytics that could disaggregate load, but their confidence in related technologies and software varied.

5. Will any baselines be appropriate for assessing demand flexibility?

The vision of continuous demand flexibility came up frequently in our interviews, with most commentary focusing on shed. Interviewees recognized that more frequent events could erode baselines, probably rendering moot assessment methodologies that draw on recent historical loads. Some interviewees suggested that measure-level assessments could help avoid this issue. More often, interviewees identified price-based approaches as alternatives to baselines for demand flexibility assessments, and a couple of people mentioned load shape-based alternatives.

Under a price-based approach, time-varying rates would signal customers to use demand flexibility to reduce costs. Because there is no settlement process (only price signals), no calculation of demand reduction is required for payments to consumers, aggregators or others. Therefore, baselines (most likely via control groups) would be required only for broader assessments. Under a load shape-based approach, customers would bid load shapes against which utilities, regional grid operators, or demand service providers could measure performance. Interviewees did not offer timelines for the deployment of these no-baseline approaches.

Some interviewees identified ways that baselines using historical day data could be improved compared to traditional practice; for example, potentially using occupancy data for M&V calculations. This view of baselines suggests that they could be more dynamic and responsive to changes in building use. Other interviewees raised the importance of grid actors, such as regional grid operators and utilities, exchanging information to avoid potentially overlapping demand flexibility events. One person identified the impact of grid actors' events on each other's baselines as an unresolved issue.

One grid operator pointed out that they operate the grid based on what loads they expect versus what loads have occurred (although historic loads inform load predictions). Thus, assessments may be based on whether controlled loads meet a certain expectation or profile, instead of how they compare with a baseline. Another grid operator indirectly referenced this potential, commenting that future management of the grid will be based on *load following generation* profiles, compared to generation following load profiles.

We also heard contrasting opinions on whether baselines for demand shed are applicable to demand shift; for example, whether you could use historical day-based demand. Modulation did not come up in most interviews. When it did, interviewees had similar responses: the short timescales made baselines difficult, but simple before-and-after measurements that do not require baseline estimates to determine impacts may be appropriate.

We found general agreement that shift and modulation baselines require more research, including at least for modulation, whether baselines are needed. In the case of batteries and distributed photovoltaics, interviewees pointed out that separate metering made baselines unnecessary because charging, discharging, and generation are directly observable.

6. What can industry, government, utilities and RTOs/ISOs do to encourage robust and cost-effective demand flexibility assessments?

Responses to this question primarily addressed the role of governments and RTOs/ISOs in enabling infrastructure investments, coordinating with utilities, and developing markets for demand flexibility. Throughout our conversations, we heard that investments in interval metering were important for the expansion of demand flexibility. Interviewees saw governments, including regulators, as having a role in enabling utilities to make these investments. One interviewee stated that not all infrastructural needs are physical, and that digital infrastructure that helps utilities manage and analyze demand flexibility resources also are important.

In our discussions, we observed a general sense that development of demand flexibility markets requires proactivity; that they will not develop without the involvement of RTOs/ISOs, governments (including regulators), utilities, and service and technology providers. Several interviewees identified a coordinating role for RTOs/ISOs and governments that responded to the needs of utilities and industry. Interviewees, for example, cited several areas where together they can standardize activities:

- Establish baseline methods
- Develop efficient settlement processes
- Adopt and promote communications standards
- Create common metrics for demand flexibility.

One individual shared that RTOs with broad geographic footprints should work with state governments and utilities to discuss development of demand flexibility. We also heard various suggestions on how RTOs/ISOs and regulators could move beyond a general coordinating role and actively shape markets for demand flexibility, including:

- Opening participation in existing markets to buildings and, in particular, batteries
- Incentivizing aggregators
- Implementing real-time prices.

Conversations on real-time prices overlapped with discussion of baselines (see responses to Question 5). We also heard that where markets for demand flexibility do exist, coordination between regional grid operators and utilities is important. One interviewee asserted that where both are operating programs, they should share data and resolve conflicts in program design.

7. What else do you think we should know about or address when writing about ex-post assessments of demand flexibility that buildings can provide?

We incorporated responses to this question into the relevant sections above.

Appendix B. Characteristics of Grid-Interactive Efficient Buildings

GEBs have smart technologies like advanced sensors and controls and data analytics that can actively manage DERs and adjust a building's load profile to co-optimize for energy costs, grid services, and occupant needs and preferences in a continuous and integrated way. Figure B-1 shows key characteristics of GEBs with demand flexibility capability.

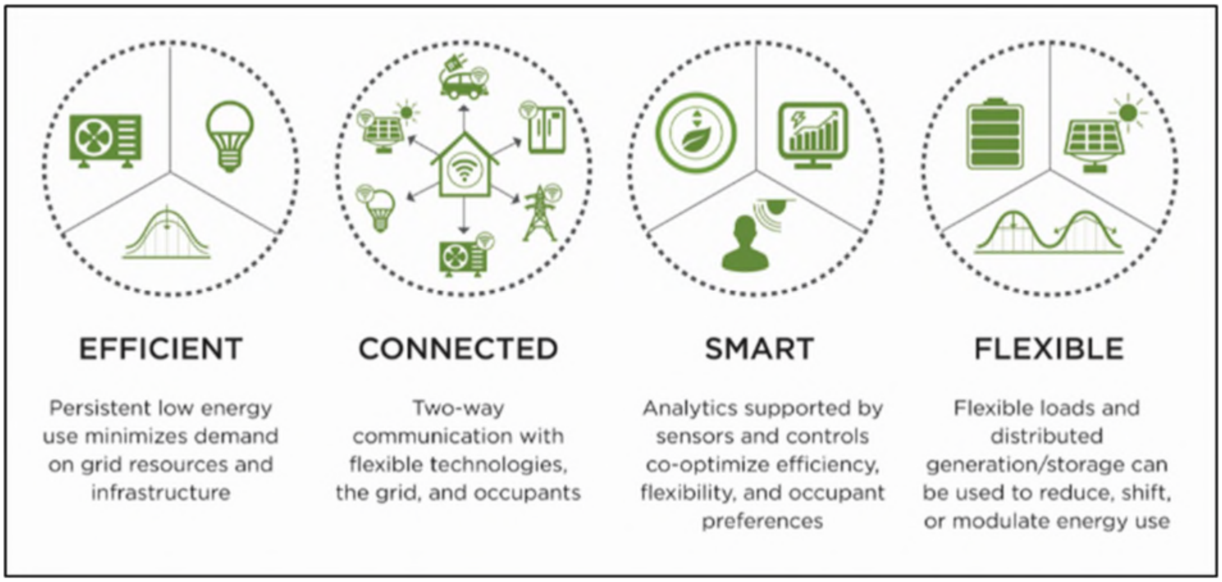


Figure B-1. Characteristics of Grid-Interactive Efficient Buildings[128]

Fortunately, two of these characteristics, connected and smart, directly support cost-effective demand flexibility assessments. Being connected implies the ability to communicate, in real time, performance information to grid operators, utilities, aggregators, and building owners and operators. Being smart implies metering and equipment status information as well as built-in analytical software that can be used to calculate (quantify) metrics of interest, which reduces the incremental cost of conducting assessments.

Following are examples of building components that can fulfill the multiple functions of: (1) control and continuous optimization of buildings, e.g., system diagnostics and fine-tuning of demand flexibility performance and (2) assessments of demand flexibility, including field validation of performance.

- Whole-building advanced meters. Such meters can provide time-differentiated energy consumption and demand data. According to a Federal Energy Regulatory Commission (FERC) staff report, "… advanced meters are now the dominant metering technology type, and the number of advanced meters in operation continues to increase in the United States…. In 2016, advanced meters accounted for approximately 47% of all residential meters, 45% of all commercial meters, and 41% of all industrial meters." By the end of 2017, almost 79 million advanced meters had been installed in the United States, representing 52% of all meters. Such meters may or may not be part of an AMI deployment.
- Smart inverter technologies. Smart inverters support distributed generation and storage technologies, providing not only the basic function of converting direct current output to alternating current, but also providing grid support functions such as voltage regulation, frequency support, ride-through capabilities for grid disturbances, and communication of electricity demand data (which can be used for assessments). For example, see IEEE 2030.5-2018 - IEEE Standard for Smart Energy Profile Application Protocol.

[128] Neukomm et al. 2019.

- Equipment and systems with monitoring capability. Information from lighting; individual motors; heating, ventilating, and cooling; and other building systems that monitor and report their status (e.g., on/off, percent of output) and power consumption (via sub-meters) can be used to assess performance.
- Control systems. Ranging from residential thermostats to commercial building management systems, control systems monitor equipment status and also may have analytical capabilities (software) that are able to execute complex strategies to adapt operation based on changing conditions over multiple timescales. Such capabilities enable calculation and reporting of energy and demand savings of the systems they operate or the building as a whole.
- Interoperable equipment, systems, and controls. These components can effectively and securely exchange data and control signals among connected devices/equipment and control systems, allowing sharing of performance (assessment) information in real time between the building and utilities or regional grid operators.
- Cybersecure equipment, systems, and controls. These components can perform the services described immediately above while maintaining end-to-end data privacy, security, and protection against unauthorized access, allowing secure communication of assessment information.
- With these components, assessments of demand flexibility may not require additional investment in metering and other equipment beyond what are already part of a GEB. However, there are two caveats:
- Commercial buildings are more likely than residential buildings to have these capabilities. If residential buildings are to provide demand flexibility beyond the commonly used load shedding and load shifting options (e.g., thermostat setbacks, precooling, air-conditioning cycling), additional hardware, software, and metering may be required for both implementation and assessment of the full range of DERs and demand flexibility modes.
- There are opportunities for improvements for analytical software, interoperability, cybersecurity, and privacy that would support demand flexibility implementation and assessments.

Appendix C. Industry-Standard Measurement and Verification (M&V) and Impact Evaluation Resources

Measurement and verification (M&V) is the typical term for performance assessments of energy efficiency and demand response activities. In the context of this report M&V is associated with the documentation of demand (and/or energy) savings or other impacts for *individual* buildings. Options for conducting M&V can involve measurements, engineering calculations, statistical analyses, and computer simulation modeling. These options are defined in the International Performance Measurement and Verification Protocol (IPMVP).[129]

Impact evaluations are similar to M&V. In the context of this report, they are defined as performance assessments of *multiple* buildings participating in a program or tariff to determine its impacts. Together, M&V and impact evaluations are known as *evaluation, measurement and verification* (EM&V).

EM&V has evolved over the past 30-plus years to provide the analytical backbone of the energy efficiency and demand response industries. The techniques developed are relevant for all DERs and demand flexibility modes that can be associated with GEBs. EM&V protocols and guidelines have been developed over time to standardize techniques and define best practices. The State and Local Energy Efficiency Action (SEE Action) Network's Evaluation, Measurement, and Verification (EM&V) Resource Portal is an EM&V resource compendium.

Some important protocols and guidelines than can be used as references for demand flexibility assessment are:

- International Performance Measurement and Verification Protocol
- Federal Energy Management Program's M&V Guidelines[130]
- American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14[131]
- Uniform Methods Protocol[132]
- FERC Measurement and Verification for Demand Response[133]
- U.S. Environmental Protection Agency's Guidebook for Energy Efficiency Evaluation, Measurement, and Verification: A Resource for State, Local, and Tribal Air & Energy Officials (2019)
- EM&V Requirements for Renewable Energy Resources drafted for the Clean Power Plan Model Rule (pages 272–283).

In addition are two recent references describing advanced M&V (also called *automated M&V* or *M&V 2.0*). These are essentially conventional M&V as it has been practiced for decades to assess energy efficiency and demand response projects and programs but with: (1) automated analytics that can provide ongoing, near-real-time savings estimates and (2) increased data granularity in terms of frequency, volume, or end-use detail.

[129] Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol (IPMVP). (multiple dates). www.evo-world.org.
[130] U.S. DOE 2015.
[131] ASHRAE 2014.
[132] Li et al. 2018.
[133] Goldberg and Agnew 2013.

- *The Status and Promise of Advanced M&V: An Overview of "M&V 2.0" Methods, Tools, and Applications*[134]

- *IPMVP's Snapshot on Advanced Measurement & Verification. Efficiency Valuation Organization*.[135]

Other important M&V resources are guidance documents used by utilities and RTOs/ISOs for assessment and settlement of their dispatchable demand response programs; for example:

- Electric Reliability Council of Texas (ERCOT)[136]
- Midwest Independent System Operation (MISO)[137]
- Independent System Operator New England (ISO-NE) Market Rule 1[138]
- PJM Regional transmission Organization[139]
- California Independent System Operator (CAISO).[140]

[134] Franconi et al. 2017.
[135] Webster 2020.
[136] ERCOT 2019.
[137] MISO 2019.
[138] ISO New England 2019.
[139] PJM 2019.
[140] California ISO 2017; California ISO 2018.

Appendix D. Demand-Side Management Strategies and Grid Services

The following table maps demand-side management strategies to the grid services they can provide. It details what the flexibility service means in the context of load changes.[141]

Note: *Response time* is the amount of time between receiving a signal from the utility/operator and the building asset responding to change the load. *Duration* is the length of time that the load change occurs.

Demand Side Management Strategies	**Grid Services**	**Description of Building Change**	**Key Characteristics**	
Efficiency	*Generation: Energy* *Generation: Capacity* *T&D: Non-Wires Solutions*	Persistent reduction in load. Interval data may be needed for M&V purposes. This is not a dispatchable service.	Duration	Continuous
			Load Change	Long term decrease
			Response Time	N/A
			Event Frequency	Lifetime of equipment
Shed Load	*Contingency Reserves*	Load reduction for a short time to make up for a sudden shortfall in generation.	Duration	Up to 1 hr
			Load Change	Short term decrease
			Response Time	<15 min
			Event Frequency	20 times per year
	Generation: Energy *Generation: Capacity* *T&D: Non-Wires Solutions*	Load reduction during peak periods in response to grid constraints or based on time-of-use (TOU) pricing structures.	Duration	2 to 4 hrs
			Load Change	Short term decrease
			Response Time	30 min to 2 hrs
			Event Frequency	<100 hrs per yr/seasonal
Shift Load	*Generation: Capacity* *T&D: Non-Wires Solutions*	Load shifting from peak to off-peak periods in response to grid constraints or based on TOU pricing structures.	Duration	2 to 4 hrs
			Load Change	Short term shift
			Response Time	<1 hour
			Event Frequency	<100 hrs per yr/seasonal
	Contingency Reserves	Load shift for a short time to make up for a sudden shortfall in generation.	Duration	Up to 1 hr
			Load Change	Short term shift
			Response Time	<15 min
			Event Frequency	20 times per year
	Avoid Renewable Curtailment	Load shifting to increase energy consumption at times of excess renewable generation output. This may be a dispatchable service but is more typically reflected through TOU pricing.	Duration	2 to 4 hrs
			Load Change	Short term shift
			Response Time	N/A
			Event Frequency	Daily

[141] Neukomm et al. 2019.

Modulate Load	*Frequency Regulation*	Load modulation in real time to closely follow grid signals. Advanced telemetry is required for output signal transmission to grid operator; must also be able to receive automatic control signal.	Duration	Seconds to minutes
			Load Change	Rapid increase/decrease
			Response Time	<1 min
			Event Frequency	Continuous
	Voltage Support		Duration	Sub-seconds to seconds
			Load Change	Rapid increase/decrease
			Response Time	Sub-seconds to seconds
			Event Frequency	Continuous
	Ramping	Load modulation to offset short term variable renewable generation output changes.	Duration	Seconds to minutes
			Load Change	Rapid increase/decrease
			Response Time	Seconds to minutes
			Event Frequency	Continuous
Generate	*Ramping*	Distributed generation of electricity to dispatch to the grid in response to grid signals. This requires a generator or battery and controls.	Duration	Seconds to minutes
			Load Change	Rapid dispatch
			Response Time	Seconds to minutes
			Event Frequency	Daily
	Generation: Energy Generation: Capacity T&D: Non-Wires Solutions		Duration	2 to 4 hrs
			Load Change	Dispatch/negative load
			Response Time	<1 hour
			Event Frequency	<100 hrs per yr/seasonal
	Generation: Energy Generation: Capacity T&D: Non-Wires Solutions	Distributed generation of electricity for use on-site and, when available, feeding excess electricity to the grid.	Duration	Entire generation period
			Load Change	Reduction/negative load
			Response Time	N/A
			Event Frequency	Daily