UC SANTA BARBARA

Participation Gap Analysis Among Energy Efficiency Programs in the Public Sector — Work Plan

Prepared by CETLab on behalf of the CAEECC Underserved Working Group

---

by Audrey Meiman, Austin Covey, Matt Lai,
Measrainsey Meng, Michelle Le, Nathaniel Villa, Sydney Litvin
under the supervision of Professor Ranjit Deshmukh

February 24th, 2021

Table of contents

I. Overview

This research study, led by the Clean Energy Transformation Lab (CETlab) at the University of California Santa Barbara, in collaboration with California Energy Efficiency Coordinating Committee (CAEECC), aims to investigate potential gaps in participation in energy efficiency programs aimed at public sector entities in California. This work plan discusses our specific research questions, methodology, data sources, and limitations.

II. Research questions

The purpose of this research is to identify gaps in participation of energy efficiency programs in the public sector. More specifically, we are interested in answering the following research questions:

1. Are there gaps in program participation by geographic areas?
2. Are there gaps in program participation by public agency criteria?
3. Are there gaps in program participation by socio-demographic groups?

Within each research question, we plan to analyze three indicators: participation, investment, and energy savings. Table 1 lists the various research questions, indicators, and measurements that we propose to analyze in this project.

Table 1. Overview of research questions, indicators, and measurements to identify underserved participants in the public sector

Research Question	Indicator	Measurement	Data Sources
Are there gaps in program participation by geographic areas?	Participation	Number of public agencies participating by county (and other unit of analysis)	CEDARS, Census Tract
	Investment	Dollars (incentives)	CEDARS, Census Tract
	Energy Savings	kWh or therms saved per public agency	CEDARS, Census Tract
Are there gaps in program participation by public agency criteria?	Participation	Number of public agencies participating by type of public agency	CEDARS
	Investment	Dollars (incentives)	CEDARS
	Energy Savings	kWh or therms saved per public agency	CEDARS

Are there gaps in program participation by socio-demographic groups?	Participation	Number of public agencies by county income level, tax revenue, and calculated CES score (pollution burden and population characteristics)	CEDARS, Census Tract, CalEnviroScreen
	Investment	Dollars (incentives)	CEDARS
	Energy Savings	kWh or therms saved per public agency	CEDARS

III. Methods

1. Preliminary exploration of participation data
    a. Classify and group programs by categories and sub-categories provided by SCE (Table 2)
    b. Group programs by spatial regions (counties)
2. Identify the overall participation, investment, and energy savings by county to evaluate whether certain areas are being less served relative to areas
3. Perform statistical and spatial analysis using information on the county's population, tax revenue, mean income, rurality, and CalEnviroscreen score to identify gaps

Table 2. Classifications of entities filing energy efficiency claims. (Source: Chris Malotte, SCE Business Plan).

State	Federal	Education	Local Governments
State Buildings, State Park Facilities, Hospitals, Correctional Facilities	Federal Buildings, US Postal Service, Hospitals, Ports, Military Bases, Tribes	Higher Education (UC/CSU/CCC)	Cities, Counties, Special Districts, Solid Waste Facilities, Water/Wastewater Facilities, Hospitals, Correctional Facilities

IV. Data Sources

The main data sources we propose to use are the California Public Utility Commission (CPUC)'s California Energy Data and Reporting System (CEDARS) database, the California Office of Environmental Health Hazard Assessment (OEHHA)'s CalEnviroScreen tool, the U.S. Census Bureau's demographics data, the California State Controller's Office, and the Index of Relative Rurality (as recommended by Lou Jacobson). Additional detail on each data source is provided below.

CEDARS

The primary data source we use is the CEDARS zip-code level dataset. The following information/variables will be used for our analysis:

- Site information: Site ID, NAICS Code, Site Zip Code, Site City
- Incentives: Total Gross Incentive
- Costs: Total Gross Measure Cost
- Savings: Total Lifecycle Gross kWh, Total Lifecycle Gross Therm

CalEnviroScreen

CalEnviroScreen is a dataset that identifies California communities that are most affected by pollution and that are often especially vulnerable to pollution's effects. The CalEnviroScreen dataset includes 20 indicators that can be grouped into four categories: exposures, environmental effects, sensitive populations, and socioeconomic factors. Indicators are provided at the census tract level. For a given census tract, each indicator has both a score and a percentile of where that census tract's indicator score falls compared to all other census tracts in the state. CalEnviroScreen (CES) also calculates a CES score for each census tract by multiplying the pollution burden and population characteristics components together. Finally, CalEnviroScreen uses this combined CES score to provide an assessment of whether the census tract is a disadvantaged community (DAC). If a census tract's CES score falls at or above the 75th percentile, it is classified as a DAC.

Because the DAC scores and percentiles are provided at the census tract level, we aggregate the data to the county level. For the CES scores, we calculate the median CES score for all census tracts within each county. We also calculate the proportion of DACs within a county by dividing the number of DAC-classified census tracts by the total number of census tracts for each county.

Census Bureau

The U.S. Census Bureau provides data at varying spatial resolutions on population, demographics, etc. For our study, we will be using county-level mean household income and population data from the Census Bureau. We will also be utilizing shapefiles from the Census Bureau's TIGER/Line database for our geospatial visualizations.

The Index of Relative Rurality

In order to account for rurality, we used the Index of Relative Rurality that came out of Purdue and Mississippi State. The IRR is a continuous, threshold-free, and unit-free measure of rurality that is spatially tracked by counties across the US. This index identifies the dimensions of rurality by population size, density, remoteness, and percentage of built-up area. For our study, we will be using the median of the IRR score from 2010 and 2000 for every individual California county.

California State Controller's Office

The California State Controller's Office (SCO) has open data on revenues, expenditures, and other financial data reported by California's counties, cities, special districts, and others. For our study, we will be using the mean tax revenues by county over the 2016 to 2018 fiscal years.

V. Data Analysis

General/broad analysis

We took a list of highlighted public sector programs provided by Chris Malotte and categorized those programs manually into four groups: (1) local government, (2) state, (3) higher education, and (4) other education. Any program that we could not categorize we kept in an "uncategorized" group. We performed some initial high-level analyses to understand the level of participation and savings in each of these categories.

Figure 1 shows how many programs fall into each category between 2017 and 2019. The local government category had the highest number of programs, followed by the uncategorized category. When comparing total gross measure costs within each category, however, Figure 2 shows that the uncategorized category does not fall far behind the local government category. That is mainly because the largest uncategorized program, Grandfathered Street Lights, has the highest gross measure costs out of any program and is significantly larger than the second highest program. Moving forward, we plan on recategorizing the Grandfathered Street Lights program as a local government program, because often municipalities are the implementers of the program. The top 10 programs with the highest gross measure costs between 2017 and 2019 are plotted in Figure 3. In terms of lifetime energy savings, the local government category also ranks highest, as shown in Figure 4.

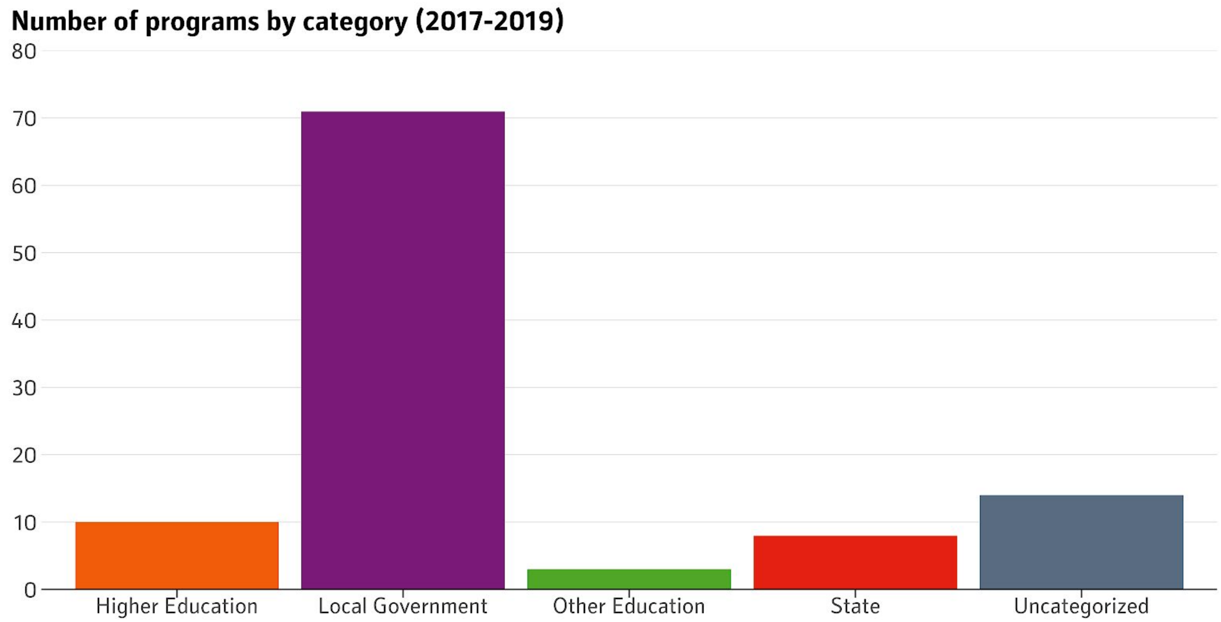

Figure 1. Number of programs by category between 2017 and 2019.

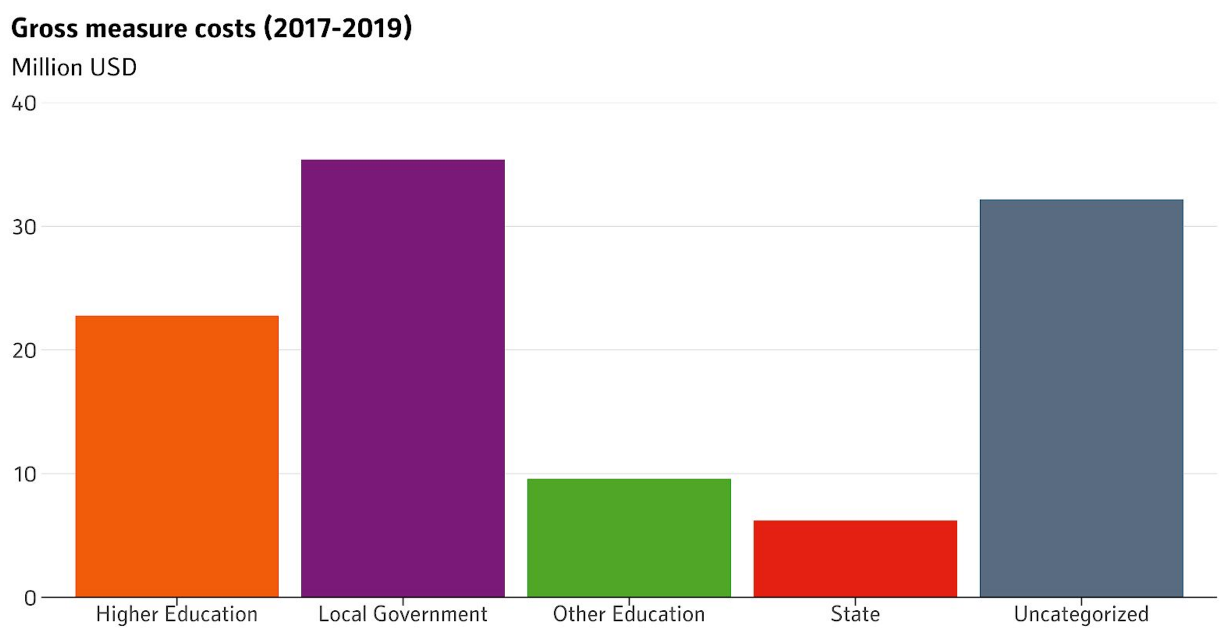

Figure 2. Total gross measure costs by category between 2017 and 2019.

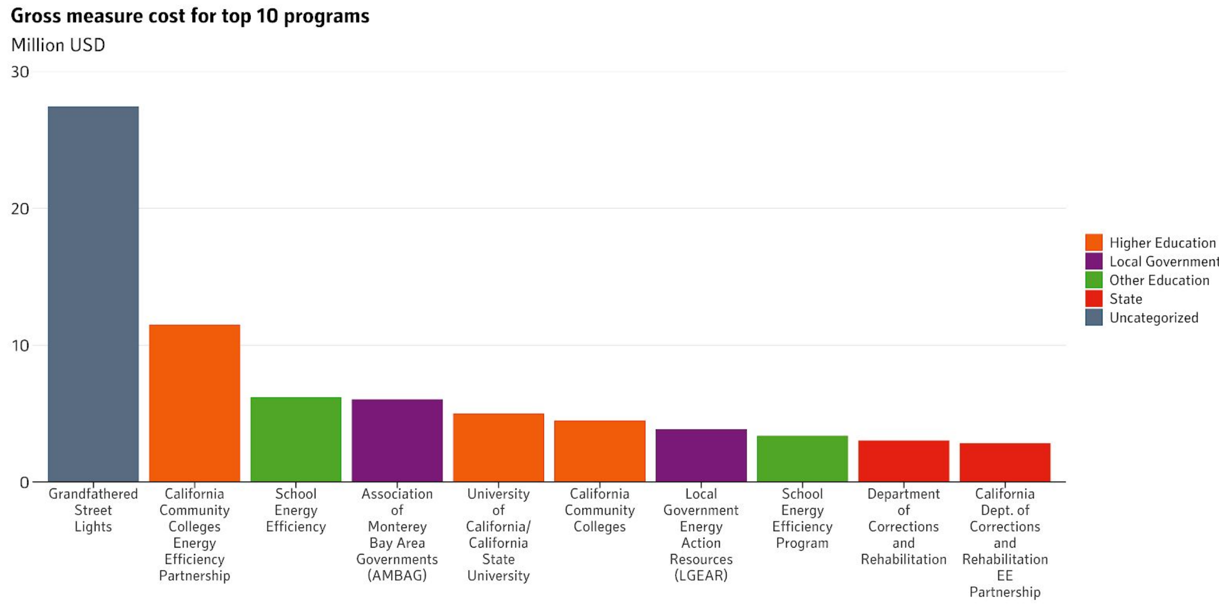

Figure 3. Gross measure costs for top 10 public sector programs. Each program's color represents the category the program belongs to.

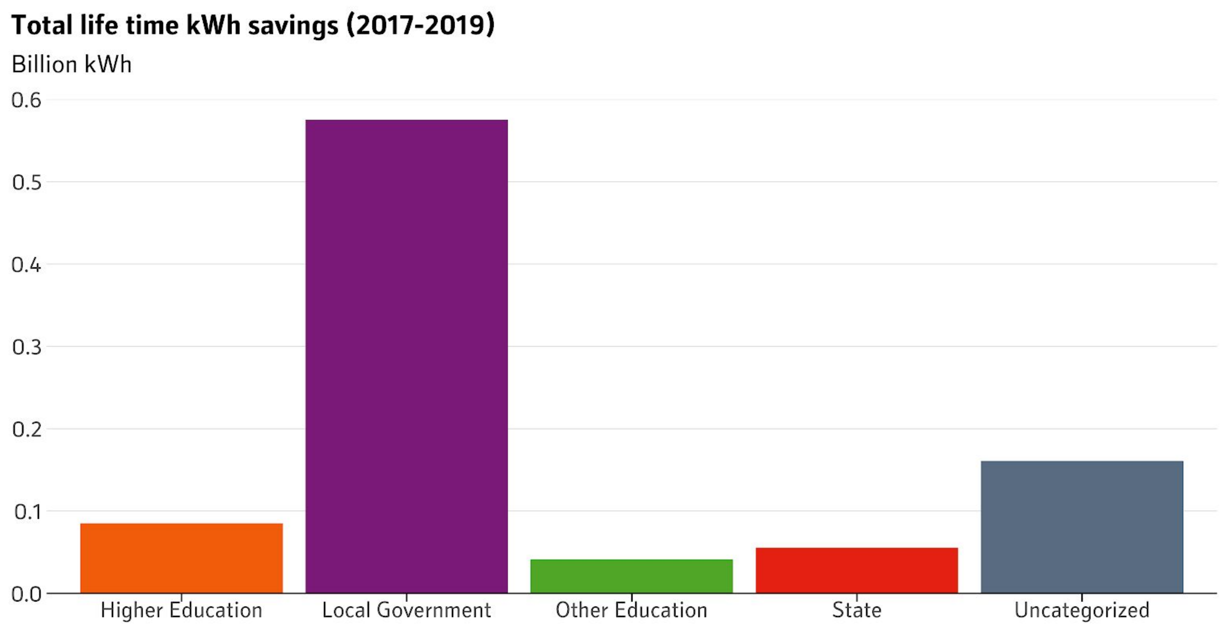

Figure 4. Total lifetime energy savings by category between 2017 and 2019.

Synthetic Variables

Along with the raw variables provided in the CEDARS and CPUC datasets, we will also be calculating additional synthetic variables (as recommended to us by the

SMB team). We believe these variables could be useful in providing further insight in the effectiveness of programs and participation in programs across categories, geographic regions, communities, etc. The synthetic variables we will be calculating are:

- Percentage of cost covered = incentive amount / measure cost
- Percentage savings or depth of savings = savings / usage
- Incentive per kWh
- Incentive per therm
- Measure cost per kWh
- Measure cost per therm

Geospatial Visualization

Because we have been provided zip code-specified claims data from the CPUC, we will be able to visualize these variables geospatially as well. Using the county TIGER/Line shapefiles provided by the Census Bureau, we will be using ArcGIS to visually analyze and project demographic information of program participants by county onto a map.

Local Governments Analysis

For the local government category, we will be performing a regression analysis to study whether participation in energy efficiency programs for these local governments are dependent on certain socio-economic characteristics. We propose to do these regressions only on local government programs because the other public agency categories (state, federal, education) have a low sample size and would likely not provide robust results. These entities are also centrally managed and their participation is likely not correlated to the characteristics of the county where they are located. We will undertake two different multivariate regression analyses: one for the cumulative budget and one for energy savings. To analyze program participation by geographic area and socio-demographic indicators, we will include these five explanatory variables in our analysis:

a. county population
b. mean household income by county
c. rurality score from the Index of Relative Rurality
d. mean tax revenue by county over the 2016 to 2018 fiscal years
e. census-tract CES score median

Our regression analysis will be performed at the county-level. The regression can be expressed as follows:

$$\text{Budget/cost; energy savings} = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Where X_1 is county population (in numbers of people), X_2 is mean household income (in units of dollars USD), X_3 is the rurality index (a value between 0-100), and X_4 is mean tax revenue (in units of dollars USD), and X_5 is the median CES score (a value between 0-100).

We will analyze the adjusted r square value, coefficients of the variables, and associated p-values to determine the effect of the variables and its significance on the budget of and energy savings from the energy efficiency programs. After estimating the coefficients, we plan to assess the accuracy and performance of our models.

VI. Limitations

Our study examines energy efficiency programs from the year of 2017 to 2019. We currently do not have access to information on program participation prior to these years. We were advised that getting this data would be difficult due to structural changes in how public utilities reported this information over the years. We will have to assume that past participation patterns are similar to what we will find for 2017 - 2019.

We were informed that there were commercial programs in which some public sector entities may have participated in. However, we have not found a way to easily identify those programs and how we could separate information for public sector participants from the larger commercial programs. Therefore, we cannot include these programs in our study and limit our scope to programs that are strictly categorized as public sector.