ParagraphS
Paragraphs Environmental Design

Original Research

DOI: 10.5281/zenodo.24156937

OPEN ACCESS

***CORRESPONDENCE**
Paulus Bawole
paulus@staff.ukdw.ac.id

RECEIVED: June 2024
ACCEPTED: August 2024
PUBLISHED: September 2024

ISSN-L 2960-2874

VOLUME 2, ISSUE 2
2024

CITATION
Bawole P. (2024). Energy Use Intensity of Public Schools Across Climate Zones: Evidence from Public Benchmarking Data. Paragraphs Environmental Design, 2(2), 10-25. DOI: 10.5281/zenodo.24156937.

COPYRIGHT
© 2024 Bawole P. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Energy Use Intensity of Public Schools Across Climate Zones: Evidence from Public Benchmarking Data

Paulus Bawole*

Faculty of Architecture and Design,
Universitas Kristen Duta Wacana, Yogyakarta

Abstract

Climate-sensitive energy benchmarks can help school systems identify unusually intensive buildings, yet differences between climate groups are easily mistaken for differences in operational efficiency. This article presents an original secondary analysis of the United States Energy Information Administration's 2012 Commercial Buildings Energy Consumption Survey, using its August 2016 public-use release and literature published before May 1, 2023. The analytical population comprises government-owned buildings classified as elementary/middle schools or high schools. Among 473 sampled buildings, 464 support comparisons across three disclosed climate groups; nine marine-climate observations remain in the national estimate but are not reported separately. Survey-weighted major-fuel site energy use intensity is estimated as a ratio of energy to floorspace, with uncertainty calculated using 197 jackknife replicate weights. Estimated intensity is 67.2 kBtu/ft²/year in very cold/cold climates, 56.8 in mixed-humid climates, and 46.8 in the combined hot-dry/mixed-dry/hot-humid group. The cold-to-combined-group difference is 20.4 kBtu/ft²/year, with an approximate 95% confidence interval of 12.8-28.0. Standardizing school subtypes reduces but does not remove the gradient. Electricity intensity moves in the opposite direction to total major-fuel intensity, while within-group distributions overlap. These findings support climate-sensitive screening but do not establish causal climate effects or identify inefficient individual schools. The article develops a benchmarking approach that distinguishes portfolio intensity from the typical building, preserves fuel-specific information, and treats educational service and indoor environmental quality as conditions of meaningful energy improvement.

Keywords: school buildings; energy use intensity; climate regions; CBECS; benchmarking

1. Introduction

Public-school energy performance matters because buildings must provide a dependable educational service while consuming resources that could otherwise support teaching, maintenance, or other public purposes. Energy use intensity, conventionally expressed as annual energy consumption divided by floor area, offers an accessible way to compare buildings of different sizes. Its apparent simplicity can nevertheless conceal differences in weather exposure, operating schedules, equipment, and the services that schools provide. A school with a higher intensity may operate longer, serve a different educational population, or condition a larger proportion of its space. The relevant research task is therefore to determine which comparisons public data can support and how those comparisons should guide investigation without becoming unsupported judgments of efficiency.

Climate provides an especially important organizing dimension because heating and cooling requirements depend on outdoor conditions. It is reasonable to expect school buildings in cold regions to incur substantial heating demand, while buildings in warmer regions may require more cooling. Annual whole-building energy, however, combines these demands with lighting, ventilation, cooking, computing, refrigeration, and other uses. Differences in fuels and equipment efficiencies further complicate the relationship between outdoor conditions and purchased energy. A useful climate comparison must consequently examine the composition of consumption as well as its annual total; otherwise, a lower total could conceal a larger electricity requirement or a different service level.

Public benchmarking information makes such comparisons possible on a scale that individual engineering audits cannot readily achieve. National surveys and local disclosure programs both provide building-level observations, but they represent different statistical populations. A probability survey seeks to characterize a wider building stock through its sampling design, whereas an administrative disclosure file reflects reporting rules, thresholds, compliance, and the information released. Treating both as interchangeable collections of schools would erase these distinctions. This article uses public benchmarking data in the broader sense of openly available building-energy evidence suitable for constructing reference distributions, and identifies the national survey basis of its empirical findings explicitly.

The empirical contribution is an independently reproducible analysis of government-owned elementary/middle and high-school buildings in the 2012 Commercial Buildings Energy Consumption Survey (CBECS). The public-use release provides a defensible combination of school classification, government ownership, annual consumption, floorspace, climate groups, and survey weights (U.S. Energy Information Administration [EIA], 2016a, 2016b). These variables do not identify school districts, individual institutions, or the quality of educational provision. The resulting estimates concern a defined stock of buildings used for schooling, rather than a census of institutions or a measure of school-system performance. This narrower formulation preserves the value of the observations while preventing the building sample from being mistaken for a school roster.

The study asks four linked questions. How does major-fuel site energy intensity differ across the reportable climate groups? How do the findings change when attention shifts from portfolio floorspace to the distribution of individual buildings? Do differences in school subtype account for the observed pattern, and what does fuel disaggregation reveal about its interpretation? These questions deliberately stop short of asking how much energy a school would save if its climate or equipment changed. The available cross-section supports descriptive comparisons and limited statistical adjustment, but it does not observe the counterfactual conditions required for those causal claims.

A second contribution is methodological: the article treats data limitations as determinants of the analysis rather than as qualifications added after results are produced. The available climate categories combine dry and humid environments, the smallest climate sample cannot sustain a separate estimate, and the newer survey release introduces a caution against climate aggregates. Such constraints can make a polished national map or an apparently precise ranking misleading.

The response here is to report only comparisons that remain interpretable after those restrictions are applied. The paper's argument is that useful school benchmarking depends as much on choosing an appropriate comparison population as on calculating a precise numerical indicator.

2. Literature and analytical framework

2.1. What school energy studies establish

School-energy research provides substantial grounds for expecting performance to vary with both environmental and organizational conditions. Dias Pereira et al. (2014) review evidence across school-building studies, making clear that school energy performance is a research field with multiple settings and measurement approaches. The range of contexts also limits the portability of a single published benchmark. A value drawn from one country's schools cannot simply be treated as an attainable norm for another stock with different climates, building forms, and schedules. The present study therefore uses the literature to identify comparison problems and plausible mechanisms, while deriving its numerical reference values from one consistently defined dataset.

School subtype is a particularly important source of heterogeneity. Hong et al. (2014) show that benchmarking comparability improves when distinctions within the school stock are examined, including differences between primary and secondary education. A single school category can therefore combine buildings with different facilities and activity patterns. Subtype adjustment alone is still incomplete: two high schools may differ in laboratories, sports facilities, meal preparation, community use, and hours of operation. The implication for this analysis is to examine elementary/middle and high-school composition explicitly, while interpreting the resulting adjustment as one controlled comparison rather than a complete explanation of energy use.

Occupancy and use also affect the meaning of an area-based benchmark. Sekki et al. (2015) examine how occupancy and space use affect energy efficiency in educational buildings, and Ouf et al. (2016) investigate occupancy-related influences in school buildings. These studies caution against assuming that floor area alone captures the service being delivered. A lightly occupied building may appear intensive per student while appearing ordinary per square metre, and a frequently used building may reverse that relationship. Accordingly, this article retains floorspace as its primary denominator because it is consistently observed, but treats educational service as an essential missing dimension in interpreting the resulting comparisons.

International evidence reinforces the need to separate metric choice from substantive performance. Daly et al. (2022) examine energy use in Australian public primary schools and identify relationships involving enrolment, floor area, climate, and remoteness. Their setting offers a close substantive comparison, but its data access conditions and national building stock differ from CBECS. Published Australian intensity levels therefore should not be used as direct targets for American school buildings. The relevant transferable finding is that a school benchmark gains explanatory value when its denominator and contextual variables are matched to the question being asked, rather than when numerical results are transferred without their original conditions.

These studies support contextual benchmarking, but they do not supply a universal adjustment formula that can be imported unchanged. Their samples, fuels, denominators, and statistical designs vary, and some datasets are unavailable for independent reanalysis. A methodological synthesis must consequently distinguish a general proposition from a locally estimated coefficient. Climate, occupancy, and subtype deserve attention across studies; the magnitude of each association requires evidence from the specific population under examination. The empirical design below follows that distinction by using prior research to justify variables and interpret limitations while calculating all reported results directly from the selected public file.

2.2. Energy intensity and efficiency

Area-based EUI is a measure of consumption density, not an independent measurement of technical efficiency. A reduction in energy per unit floor area may reflect better equipment or control, but it can also reflect fewer operating hours, less conditioned space, or lower occupancy. The energy numerator cannot distinguish these

pathways by itself. This creates a basic interpretive requirement: a benchmark can identify buildings whose consumption deserves explanation, while an efficiency claim requires additional evidence about the service maintained. In school applications, that service includes usable teaching space and acceptable indoor conditions, making a low EUI an incomplete success criterion.

Site and source energy answer different accounting questions. Site measures describe energy delivered or used at the building boundary, whereas source accounting incorporates upstream conversion and delivery factors. A building using electricity and one using fuel for heating can therefore change their relative position when the accounting boundary changes. EPA's school scoring approach uses source energy and adjusts for specified operating and weather characteristics (U.S. Environmental Protection Agency [EPA], 2018). This article instead analyzes the major-fuel site measure available in CBECS. Its results are appropriate for understanding that consumption boundary and cannot be substituted for ENERGY STAR scores, greenhouse-gas intensities, or lifecycle energy assessments.

Weather normalization is frequently treated as though it makes buildings in different regions directly comparable. EPA (2020) distinguishes adjustment to a building's normal local weather from adjustment across climatic conditions. The distinction matters because replacing an unusually cold year with a location's typical year still leaves a cold location colder than a warm one. A weather-normalized EUI can improve year-to-year comparison without establishing a common national climate baseline. For the present study, the observed annual values are therefore retained as historical consumption measures, and the climate-group comparisons are described as associations in that year rather than as fully normalized performance rankings.

The denominator introduces a separate distinction between the intensity of a portfolio and that of a typical building. Dividing weighted total energy by weighted total area gives greater influence to buildings that account for more floorspace. Averaging each building's EUI using survey weights gives influence according to its representation in the building population, irrespective of its size. Neither quantity is inherently preferable, but they answer different questions and may produce different numerical results. Reporting the portfolio ratio alongside building-level quantiles makes that distinction visible and prevents a national floorspace benchmark from being presented as the median school-building experience.

2.3. Climate as exposure and context

Climate categories summarize environmental conditions, but each category also contains a historically developed stock of buildings and systems. Construction practices, heating fuels, cooling availability, and patterns of school use can vary geographically alongside climate. An observed between-group difference consequently combines climatic exposure with features of the buildings and institutions located there. Statistical adjustment can describe how the comparison changes when selected characteristics are held constant, but it cannot guarantee that every relevant difference has been removed. The appropriate interpretation is a conditional association within the observed stock, with causal mechanisms examined as hypotheses rather than established effects.

Heating and cooling create a plausible physical account of the expected gradient, although that account does not imply that total EUI must rise monotonically with either degree-day measure. Internal gains can offset heating needs while increasing cooling demand, school calendars alter exposure to summer weather, and equipment converts thermal demand into purchased energy with different efficiencies. A school that uses efficient electric heating may also have a different fuel pattern from a nearby school with a combustion system. These interactions justify examining electricity and natural gas separately, while cautioning against reading a fuel-specific difference as a direct measurement of a particular end use.

Indoor environmental quality provides an essential boundary on interpretations of savings. Fisk (2017) reviews evidence concerning ventilation in schools, while Haverinen-Shaughnessy and Shaughnessy (2015) examine associations of classroom ventilation and temperature with academic achievement. These

sources support taking the learning environment seriously, but they do not allow achievement effects to be inferred from an annual energy ratio. A high EUI does not demonstrate good ventilation, and a low EUI does not demonstrate poor ventilation. The practical synthesis is to pair energy screening with evidence on indoor conditions before identifying a reduction as an improvement in school performance.

The framework therefore distinguishes three objects of comparison: the historical energy burden of the building stock, the position of an individual building within a peer distribution, and the energy needed to deliver a specified service. Public annual records are relatively well suited to the first object and can support cautious screening for the second. They are less complete for the third because service quality, detailed schedules, and system operation are only partially observed.

Preserving these distinctions allows the same dataset to be useful without requiring it to answer questions beyond its measurement design.

3. Data and methods

3.1. Source selection and historical scope

The primary data are the August 2016 public-use release of the 2012 CBECS. The file contains 6,720 responding commercial-building records and supplies the variables necessary for a reproducible school-building domain analysis (EIA, 2016a). Its age is a limitation for describing a later building stock, but the fixed release has two advantages for this study: it predates the literature cutoff and its relevant climate classifications are available for every selected record. The analysis therefore addresses a historical baseline, not current performance. Later technological changes, retrofits, and changes in educational operations are outside the period represented by the observations.

Selecting the older release is also a response to a specific disclosure constraint. The January 2023 guide to the 2018 public-use file recommends against producing climate-zone estimates for any aggregate because climate information is suppressed for some cases, and its climate grouping differs from the earlier survey (EIA, 2023). The 2012 file is consequently not pooled with 2018 to create a spurious time series. Nor is the newer file treated as automatically preferable merely because it has a later reference year. The study instead uses the earlier internally consistent classification and limits its interpretation to the population and period that classification supports.

The literature cutoff is April 30, 2023, inclusive. It restricts the documentary basis of the paper rather than implying that its calculations were undertaken in 2023. The analysis is an original reanalysis of a previously released historical file, and the accompanying materials preserve its download provenance and checksum. This distinction matters because a newly calculated statistic can be based entirely on older evidence without being an older published result. The article accordingly identifies its numerical estimates as author calculations and cites the underlying data release, avoiding any suggestion that EIA previously published these particular school-climate tables.

3.2. Population and observation unit

The analytical domain is defined by two explicit filters: PBAPLUS equals 28, elementary/middle school, or 29, high school; and GOVOWN equals 1, government ownership. This produces 473 records. Ownership is preferable to inferring public status from a building's name, but it is not equivalent to a complete administrative definition of public-school operation. A publicly operated school in privately owned premises could fall outside this domain, and ownership alone does not establish district governance or funding arrangements. Throughout the analysis, “public schools” therefore refers operationally to these government-owned school buildings, with that boundary retained in interpreting every population estimate.

The unit is a building, which is consequential when interpreting counts and weights. One educational institution may occupy several buildings, while a campus energy meter or local benchmarking property may cover more than one structure. CBECS building weights cannot be interpreted as counts of separate schools, districts, or students. Conversely, an institutional roster cannot be substituted for building weights without a verified linkage and consistent physical boundaries. The study reports sampled buildings and weighted building-stock quantities separately, preventing a numerically large population estimate

from implying that the survey identified an equivalent number of educational institutions.

All selected records have positive floorspace, positive major-fuel consumption, and positive final survey weights. None lacks a disclosed climate category, so the main analysis does not delete school records because their climate or outcome is missing. This completeness does not mean that every underlying value was directly measured: the released file includes imputation and masking. Those features are part of the public-use evidence rather than problems that can be removed by silently discarding inconvenient cases. The primary estimates therefore retain the released domain, and their interpretation distinguishes sampling precision from the broader uncertainty associated with measurement and processing.

3.3. Climate groups and measurement boundary

The Building America climate-region classification, as combined for the 2012 public-use file, permits three sufficiently populated comparisons: very cold/cold, mixed-humid, and the combined hot-dry/mixed-dry/hot-humid group. Their unweighted sample sizes are 178, 144, and 142, respectively. These categories are not interchangeable with numerical ASHRAE or IECC zones. The merged third group must retain its full identity because its dry and humid environments cannot be separated in the public file. It is used here as a disclosed comparison group, with shortened references to the “combined group” only after the full category has been specified.

Nine additional observations belong to the marine category. They remain in the national domain, but a separate marine estimate is withheld because the sample falls below the 20-observation publication threshold described in the 2012 guide; estimates with relative standard errors above 50% are also withheld (EIA, 2016b). This creates a deliberate asymmetry between the national total and the three reported climate groups. A fourth point on a chart might look more complete, yet it would carry an evidentiary weakness hidden by that symmetry. Suppression is therefore part of the analytical result: the released sample supports three climate comparisons, not a comprehensive ranking of every climate category.

The main outcome is major-fuel site EUI, calculated as MFBTU divided by SQFT for each building. MFBTU expresses annual consumption in thousand British thermal units and combines electricity, natural gas, fuel oil, and district heat. It is not a comprehensive account of every possible energy carrier, upstream energy, or emissions. For that reason, “total” in discussion of the empirical tables means the sum of these major fuels, and the more precise label is retained in captions. Values are presented in kBtu/ft²/year; multiplying by approximately 3.1546 converts them to kWh/m²/year without changing the accounting boundary.

Fuel disaggregation uses the corresponding electricity and natural-gas energy variables with the same full-domain floorspace denominator. This choice yields additive contributions to major-fuel intensity and allows meaningful comparison of the fuel mix. It differs from reporting the intensity only among buildings that use a particular fuel, a statistic that can change merely because fuel adoption differs across regions. Missing entries for a non-used fuel contribute zero after their non-use meaning is checked. The denominator is kept fixed so that the fuel results explain the portfolio total rather than describe differently selected populations.

3.4. Estimands and uncertainty

For climate group g, the primary portfolio intensity is $R_g = \Sigma(w_i E_i)/\Sigma(w_i A_i)$, where E_i is annual major-fuel energy, A_i is floorspace, w_i is FINALWT, and both sums run over the group's school buildings. The weights address the survey's unequal selection probabilities and participation adjustments; floorspace weighting then determines the portfolio intensity. This is distinct from treating the 473 buildings as a simple random sample or treating all sampled observations equally. The estimator follows the ratio-of-totals structure illustrated by EIA, while applying it to a school-ownership domain defined specifically for this article.

Building-level distributions are estimated using FINALWT applied to each building's E_i/A_i. The weighted median and quartiles use the inverse weighted empirical cumulative distribution, with no interpolation between observed values. These measures describe the distribution of represented

buildings rather than the distribution of floor area. They do not carry confidence intervals in this article, and their observed overlap is not used as a test of equality between climate populations. Their purpose is complementary: they reveal how much variation a single group-level ratio can conceal when it is used to characterize an individual building.

Uncertainty for ratios and smooth statistics is calculated with all 197 supplied jackknife replicate weights. The statistic is recomputed under each replicate, and its variance is the sum of squared deviations from the full-sample estimate; the replicate coefficient is one for this release (EIA, 2016b). Approximate 95% intervals use the estimate plus or minus 1.96 standard errors. These intervals describe sampling uncertainty under the released design and are not predictive ranges for individual buildings. They also do not encompass every source of bias, making a narrow interval insufficient on its own to establish measurement validity or causality.

Between-group differences are recomputed within each replicate before their standard errors are calculated. This retains covariance induced by the common survey design instead of combining independently calculated standard errors as if the groups came from unrelated surveys. The three pairwise intervals are descriptive and are not adjusted for multiple comparisons. The main substantive emphasis is the magnitude and interpretation of the gradient, rather than a search for nominal statistical significance. This approach gives the reader a transparent uncertainty calculation while avoiding an inflated claim that several interval comparisons independently prove a climatic mechanism.

3.5. Composition and robustness checks

The first adjustment standardizes each climate group to a common school-subtype composition. Separate portfolio ratios are calculated for elementary/middle and high-school buildings within each group, then combined using their national weighted floorspace shares. Those shares are 66.7% and 33.3%, respectively, and they are recalculated within every replicate when uncertainty is estimated. The procedure asks what the group ratios would be under a common subtype allocation of area. It does not create a counterfactual stock with identical buildings, but it directly tests whether subtype composition alone accounts for the unadjusted ordering.

An exploratory weighted log-linear model provides a second, more conditional comparison. It relates log EUI to climate indicators, log floorspace, school subtype, weekly operating hours, months in use, cooking activity, and the percentages of the building heated and cooled. These predictors are specified to examine observed service and building differences, rather than selected through a search for significant coefficients. Conditioning on heated and cooled space can remove part of the pathway by which climate affects energy demand. Accordingly, the model estimates adjusted associations within the observed stock and is not interpreted as a causal climate effect or as an individual-building efficiency score.

The model uses the 464 buildings in the reportable climate groups, with very cold/cold as the reference category. Eleven inapplicable heating fractions and 19 inapplicable cooling fractions are coded as zero only where the corresponding no-use indicators confirm that the space is unheated or uncooled. This preserves structural non-use without treating an unknown value as zero. The complete model is refitted under each of the 197 replicate weights, and coefficient covariance is the sum of the replicate-deviation outer products. These procedures make the conditional estimates reproducible while retaining their distinction from the portfolio ratios.

Influence is assessed by removing each sampled building in turn while holding the released final weights fixed. The resulting range indicates whether a single observation drives the principal portfolio ratio. It is not a substitute for the jackknife variance calculation, because deleting one record in this way does not reproduce the survey design. The primary estimates retain every valid record, including unusually high intensities. This combination preserves genuine heterogeneity while showing the reader whether a result depends on one extreme observation or on a more broadly distributed pattern.

Computational validation uses an example outside the school domain to check the implementation independently of the study's

substantive expectations. For all commercial buildings in the 10,001-25,000 ft² category, the code reproduces EIA's published ratio of 62.103105 kBtu/ft² and standard error of 3.144442 to the displayed precision. Agreement with that example does not validate every substantive assumption or transform imputed data into direct measurements. It does establish that the weight handling, unit interpretation, and replicate calculation behave as documented. The analysis materials therefore pair numerical reproducibility with an explicit account of the remaining interpretive limits.

4. Results

4.1. Population scale and portfolio intensity

The 473 sampled records represent an estimated 194,485 government-owned school buildings, occupying approximately 8.09 billion ft² and consuming 475.2 trillion Btu of major-fuel energy in 2012. These quantities describe the weighted building stock; they do not imply that the United States contained 194,485 separate public schools. Estimated building counts are also less precise than the portfolio EUI, with a relative standard error of 11.7% for the count compared with 3.0% for intensity. The ratio benefits from the joint relationship between its numerator and denominator, so its precision should not be transferred to other population quantities without their own uncertainty estimates.

National major-fuel site EUI is 58.7 kBtu/ft²/year, with an approximate 95% confidence interval of 55.3-62.2. In SI units, the point estimate is approximately 185.2 kWh/m²/year. This is a historical weighted portfolio ratio, not a recommended operating target or a current national standard. It also includes the marine observations that are omitted from separate climate reporting. The estimate is most useful as a reference for the selected building population and accounting boundary; interpreting a particular school's position requires the climate, distributional, and service distinctions examined below.

The reportable climate groups exhibit an ordered portfolio gradient: 67.2 kBtu/ft²/year in very cold/cold climates, 56.8 in mixed-humid climates, and 46.8 in the combined dry/hot-humid group. Their approximate 95% intervals are 61.0-73.4, 51.3-62.3, and 42.4-51.1, respectively. Table 1 makes both the estimates and their sample support visible. These differences are appreciable, but they describe the buildings found in each group under observed 2012 conditions. They do not isolate outdoor climate from the fuels, systems, construction patterns, or educational uses associated with those locations.

Table 1. Major-fuel site energy intensity and building distributions

Climate region	n	Portfolio EUI	95% CI	Building median	Building IQR
Very cold/cold	178	67.2	61.0-73.4	64.3	46.6-94.2
Mixed-humid	144	56.8	51.3-62.3	52.4	37.6-78.5
Hot-dry/mixed-dry/hot-humid	142	46.8	42.4-51.1	34.7	25.0-52.0
All climate regions	473	58.7	55.3-62.2	47.5	28.3-74.8

Note. EUI is in kBtu/ft²/year. Portfolio EUI is weighted total energy divided by weighted total floorspace. Confidence intervals are approximate 95% normal intervals based on 197 jackknife replicates. Medians and interquartile ranges (IQRs) use building survey weights; IQRs are not confidence intervals. All regions includes nine marine observations whose separate estimates are withheld. Source: Author calculations from EIA (2016a).

The cold-to-combined-group contrast is 20.4 kBtu/ft²/year, with a design-based standard error of 3.9 and an approximate 95% interval of 12.8-28.0. Relative to the combined group's point estimate, the cold-group ratio is about 43.7% higher. That percentage expresses a cross-sectional comparison, not the savings available from a retrofit or the penalty imposed by climate alone. The contrast is nevertheless large enough to matter for screening: applying one

unadjusted national threshold would systematically evaluate these historical group distributions against different starting points.

The adjacent contrasts are smaller but remain informative. Very cold/cold minus mixed-humid is 10.4 kBtu/ft²/year, with an approximate interval of 2.2-18.7; mixed-humid minus the combined group is 10.0, with an interval of 3.0-16.9. The group-specific intervals overlap in places, which does not contradict the intervals calculated directly for their differences. Each contrast incorporates the survey covariance and answers a different uncertainty question. The appropriate conclusion is evidence of a historical group gradient under this estimator, while avoiding the stronger claim that every school in a colder category consumes more than every school elsewhere.

4.2. Building distributions and fuel composition

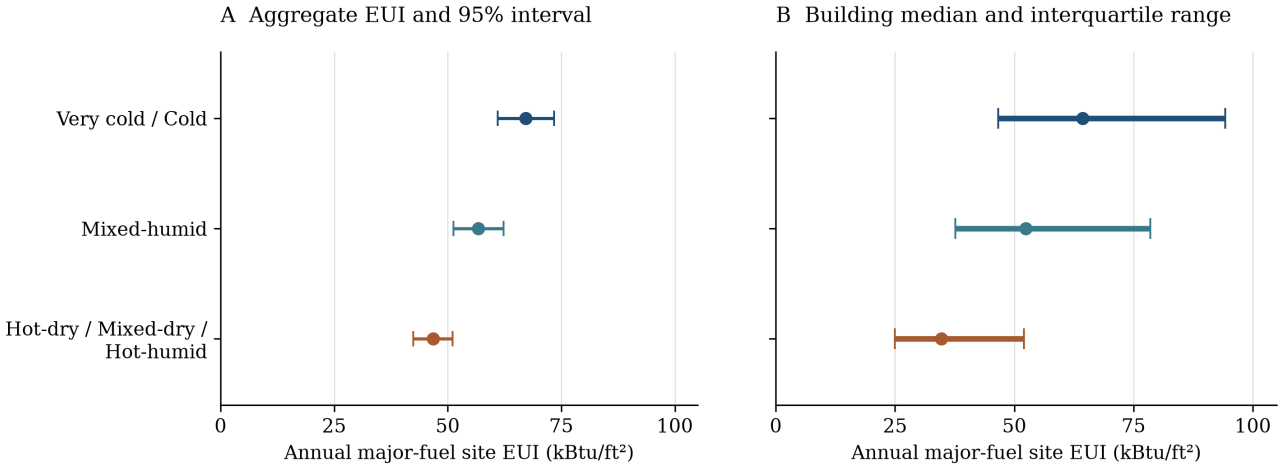


Figure 1. Portfolio uncertainty and building-level dispersion across climate regions

Note. Panel A shows ratios of survey-weighted energy to survey-weighted floorspace, with approximate 95% confidence intervals from 197 jackknife replicates. Panel B shows building-weighted medians and interquartile ranges, which describe dispersion rather than estimation uncertainty. Separate marine estimates are withheld. Source: Author calculations from EIA (2016a).

The building-weighted medians are 64.3, 52.4, and 34.7 kBtu/ft²/year in the same climate-group order. Their interquartile ranges are 46.6-94.2, 37.6-78.5, and 25.0-52.0. These ranges overlap, and each group's interquartile width exceeds the 20.4-unit difference between the cold and combined portfolio ratios. Quartile widths and portfolio differences are distinct descriptive statistics, so this comparison is not a formal variance decomposition. It does show why climate membership is insufficient for assigning an individual building's performance: substantial heterogeneity remains within every reportable category.

The distinction between portfolio and building estimands is especially visible in the combined group, where the portfolio ratio is 46.8 but the building median is 34.7. A median gives the middle of the represented building distribution, while the portfolio ratio reflects where energy and floorspace are concentrated. Neither statistic should be relabelled as the other or treated as a computational inconsistency. For a district planning total expenditure or energy reduction, the portfolio measure is directly relevant; for identifying an unusual building, its position within an appropriate building distribution provides different information.

Electricity intensity reverses the ordering observed for major-fuel EUI. Its portfolio contribution is 30.1 kBtu/ft²/year in very cold/cold climates, 34.8 in mixed-humid climates, and 37.8 in the combined group. Natural-gas intensity moves in the opposite direction, from 33.6 through 17.7 to 8.9. Table 2 shows the separate quantities using a common denominator. Thus, a conclusion that the combined group is universally less energy intensive would be too broad: it has lower major-fuel intensity and higher electricity

intensity, illustrating why the fuel boundary is part of the substantive finding.

Table 2. Fuel-specific contributions to portfolio energy intensity

Climate region	Electricity EUI (SE)	Natural-gas EUI (SE)
Very cold/cold	30.1 (1.7)	33.6 (2.8)
Mixed-humid	34.8 (2.4)	17.7 (1.5)
Hot-dry/mixed-dry/hot-humid	37.8 (2.3)	8.9 (1.3)
All climate regions	33.6 (1.2)	22.1 (1.3)

Note. All values are kBtu/ft²/year; standard errors (SEs) use 197 jackknife replicates. Both fuels use the same full-domain floorspace denominator, including buildings not using that fuel. Electricity and natural gas do not exhaust major-fuel EUI; fuel oil and district heat are included in Table 1 totals. Sparse fuel breakdowns are not reported. Source: Author calculations from EIA (2016a).

The cold-to-combined natural-gas difference is approximately 24.7 kBtu/ft²/year, while the corresponding electricity difference is approximately -7.7. These opposite movements partly offset one another in the total. Natural gas cannot be equated with heating alone, and electricity cannot be equated with cooling alone; both serve multiple activities in school buildings. The decomposition is consistent with differences in fuel use and end-use demand, without identifying their causes. Its practical implication is that a total-EUI comparison should be accompanied by fuel-specific investigation before an equipment or operational intervention is selected.

4.3. Composition and conditional comparisons

The observed groups differ in more than climate. High-school buildings represent approximately 31.5% of weighted building counts in the cold group, 18.3% in mixed-humid climates, and 13.3% in the combined group. Weighted mean building sizes also differ substantially, at approximately 59,526, 79,664, and 18,659 ft². These are contextual point estimates rather than causal explanations of the EUI gradient. Their variation establishes a concrete reason to avoid interpreting the unadjusted comparison as if it involved otherwise identical school buildings placed in different environments.

Applying common national subtype floorspace shares produces standardized portfolio intensities of 65.5, 56.8, and 48.0 kBtu/ft²/year. The cold-to-combined point difference consequently falls from 20.4 to 17.5, an attenuation of about 14.1%, while the group ordering remains. The adjustment controls only the mix of elementary/middle and high-school floorspace, leaving within-subtype variation untouched. It therefore weakens an explanation based entirely on subtype composition without demonstrating that the remaining difference is caused by climate. The unadjusted and standardized quantities together give a more defensible account than either alone.

The exploratory regression includes all 464 observations in the three reportable groups. No additional records are lost to incomplete predictors after explicitly inapplicable heating and cooling shares are coded as structural zeros. Relative to the cold group, the adjusted geometric EUI ratio is 0.885 for mixed-humid buildings, with an approximate 95% interval of 0.715-1.096, and 0.688 for the combined group, with an interval of 0.511-0.927. These are exponentiated log-model coefficients, not adjusted versions of the arithmetic portfolio ratios. Their interpretation is conditional on the listed building and service characteristics and the model's functional form.

Table 3. Exploratory survey-weighted regression of log EUI

Predictor	Log-scale coefficient	SE	Exponentiated estimate (95% CI)
Intercept	3.116	0.746	-
Mixed-humid climate	-0.122	0.109	0.885 (0.715-1.096)
Combined dry/hot-humid climate	-0.374	0.152	0.688 (0.511-0.927)
ln(floorspace / 50,000 ft²)	-0.062	0.057	0.940 (0.840-1.052)
High-school building	0.133	0.095	1.143 (0.948-1.377)
Weekly hours / 10	0.040	0.023	1.041 (0.994-1.090)
Months in use	-0.007	0.050	0.993 (0.900-1.095)
Cooking activity	0.273	0.175	1.313 (0.932-1.851)
Heated fraction (0-1)	0.598	0.213	1.819 (1.199-2.760)
Cooled fraction (0-1)	0.063	0.142	1.065 (0.806-1.407)

Note. N = 464; very cold/cold is the reference climate. The combined group is hot-dry/mixed-dry/hot-humid. Weighted R^2 = 0.203. Standard errors use complete refits under all 197 jackknife replicate weights. Exponentiated coefficients are conditional geometric EUI ratios per stated predictor increment, not arithmetic portfolio ratios or causal effects. School subtype and cooking are binary indicators. Source: Author calculations from EIA (2016a).

The regression's weighted R^2 is 0.203, indicating that the specified predictors explain only part of the observed variation in log EUI. That remaining variation cannot be labelled waste: it includes omitted characteristics, measurement differences, model limitations, and potentially operational performance. The mixed-humid interval also includes one, whereas the unadjusted arithmetic contrast with the cold group excludes zero. This is a useful reminder that changing the estimand and conditioning set changes the question. The regression supplements the descriptive pattern but does not justify collapsing its uncertainty into a categorical ranking of climate groups or individual schools.

4.4. Robustness and evidentiary limits

The single-record influence check does not indicate that one extreme observation determines the main gradient. Removing one building at a time gives cold-group portfolio ratios between 66.0 and 68.1, mixed-humid ratios between 56.1 and 57.2, and combined-group ratios between 45.8 and 47.5. These ranges are small relative to the cold-to-combined difference, and the ordering persists throughout the check. They are not confidence intervals and do not address coordinated measurement errors, imputation assumptions, or unobserved selection. The result supports resistance to a single-record perturbation while leaving those broader limitations intact.

The findings are consequently stronger for population description than for diagnosis. They establish a weighted historical gradient, demonstrate overlap within climate groups, and show that fuel-specific rankings differ from total-energy rankings. They do not identify the school buildings that could reduce consumption while maintaining the same educational service, nor do they quantify the savings from any particular measure. This separation is central to the contribution: the public file can establish where a benchmark needs context, even when it cannot establish why every building occupies its observed position.

5. Discussion

5.1. Interpreting the climatic gradient

The higher major-fuel EUI of cold-region school buildings is consistent with a substantial heating burden, especially given the natural-gas

pattern. Physical plausibility strengthens that interpretation, but plausibility does not eliminate geographic confounding. The cold group also differs in school subtype and building size, while other relevant differences remain unmeasured or only coarsely represented. The defensible synthesis is therefore that heating-related conditions are a credible contributor to the observed gradient, with its exact causal share unresolved. Presenting the entire 20.4-unit difference as a pure climate effect would exceed what the design can identify.

The fuel reversal clarifies why "better performance" must be tied to a specified objective. If the concern is major-fuel energy per unit area, the combined group has the lower ratio; if it is electricity use per unit area, the cold group has the lower ratio. Neither finding alone determines carbon emissions, utility expenditure, or peak electrical demand. Those outcomes depend on additional factors that are not constant across fuels, regions, or time. The study therefore supports a set of complementary indicators, each with a declared purpose, rather than a single EUI ranking intended to settle every energy-policy question.

The difference between adjustment and explanation is equally important. Subtype standardization changes the cold-to-combined comparison moderately, and the exploratory model retains a lower conditional geometric intensity in the combined group. It would be tempting to attribute the unexplained remainder to climate after "controlling for" the included variables. That interpretation would require stronger assumptions about omitted variables, functional form, and the causal position of heating and cooling coverage. The adjustments instead provide a structured sensitivity analysis: the pattern is not solely a consequence of the measured subtype mix, but its causal decomposition remains open.

The findings also illuminate a tension between broad comparisons and locally meaningful peers. A sufficiently populated national climate category offers support for stable estimation, yet it combines school buildings with different systems and services. A narrowly defined local peer group may offer better operational similarity but too few observations or too much shared institutional history for reliable generalization. Papadopoulos and Kontokosta (2019) demonstrate the relevance of context-sensitive grading in a different building sector, without establishing a school-specific solution. For public schools, a reasonable approach is to use national groups for orientation and refine comparisons with verified local information before making consequential decisions.

5.2. What public evidence can support

Open data strengthens research by allowing definitions, estimates, and computational choices to be checked. The reproducible CBECS analysis shows that access to a public file can support more than a literature-based argument: it can produce original, uncertainty-qualified evidence. Openness does not remove the effects of masking, imputation, or limited variable detail, however. The public-use product is a deliberately transformed representation of the underlying survey. Its value lies in allowing transparent analysis within that representation, rather than in licensing reconstruction of suppressed geography or claims of precision beyond the released design.

Municipal disclosure data can offer more recent and locally actionable information than a national survey. Their usefulness for school comparisons nevertheless depends on whether public-school properties are covered, whether reporting is complete, and whether campus and building boundaries match the energy meters. Palmer and Walls (2017) identify data-quality and evaluation problems in benchmarking and disclosure programs more broadly. Those problems do not invalidate administrative data, but they make naive pooling inappropriate. A cross-city school analysis should therefore begin with a coverage and boundary audit before calculating a climate coefficient from whatever records happen to be available.

Predictive success within disclosed buildings also differs from representativeness across a wider stock. Kontokosta and Tull (2017) show that city-scale prediction requires attention to the distinction between covered buildings and the population to which predictions are extended. Schools excluded by size thresholds or missing reporting may differ systematically from those observed. A precise regression fitted to reporting

properties can therefore remain a poor estimate of a national climate relationship. The national survey used here addresses this problem through its sampling design, while accepting a different limitation: a historical reference year and restricted geographic detail.

The article's contribution should not be confused with a systematic review or a claim to have discovered the role of climate in school energy use. Prior school research already establishes the importance of contextual variation. What is added here is a transparent combination of ownership filtering, explicit estimands, replicate-weight uncertainty, subtype standardization, and fuel decomposition for a particular public-use dataset. Each component addresses a specific inferential failure that a simple ranking could conceal. The resulting evidence is valuable because a reader can reproduce both the numerical gradient and the limits on its interpretation.

5.3. Educational service and the meaning of low consumption

A school-energy benchmark becomes more useful when it helps facilities staff investigate high consumption while preserving educational conditions. It becomes less useful when it rewards reductions whose source is unknown. Borgstein et al. (2016) emphasize service provision and comfort in evaluating non-domestic energy performance, a perspective particularly relevant to teaching environments. This does not imply that low-EUI schools are providing inadequate service; the dataset contains no basis for that generalization. It implies that confirmation of service quality is a necessary second step before low consumption is adopted as a model for other schools.

Ventilation illustrates why equipment replacement and operational performance should be assessed separately. Chan et al. (2020) document ventilation shortcomings in recently retrofitted California classrooms, showing that new HVAC equipment alone does not guarantee the intended indoor outcome. Their findings cannot establish the prevalence of such problems in the national CBECS sample. They do support an operational lesson: an energy project should be evaluated through measured delivery and commissioning, rather than through equipment age or energy reduction alone. For benchmarking, this means that an apparently improved EUI should trigger verification of the conditions under which the improvement occurred.

Area-normalized intensity can also interact with demographic change in ways that complicate school-system decisions. A building may retain nearly the same heating and lighting requirements after enrolment falls, increasing energy per student without a comparable change in EUI. Conversely, accommodating more students in existing space can improve energy per student while increasing ventilation or cooling demand per unit area. Neither metric fully describes the trade-off. A school system should therefore preserve separate records of energy, floor area, enrolment, and use schedules, allowing the causes of a change to be examined instead of compressing them prematurely into a single ratio.

6. Implications for benchmarking and policy

The results support a staged approach to public-school benchmarking in which data definition precedes performance judgment. Each record should first establish the physical boundary, ownership or governance classification, reporting year, floor-area convention, fuels included, and treatment of missing consumption. Collecting these fields adds work, and a district may initially have incomplete records. The immediate response should be to identify the missing boundary information and qualify the comparison, rather than to fill gaps with an assumed national average. This prioritizes a usable baseline whose meaning remains stable when future reports are added.

The second stage should report the portfolio ratio and a building-level distribution together. A district concerned with its total energy burden needs the former, while staff selecting buildings for investigation need the latter. Using both avoids treating the median as a budget forecast or interpreting the portfolio average as the experience of a typical facility. Climate and school subtype should enter this screening where sample support permits. The historical values in this article demonstrate the consequences of that choice, but they should be updated with

appropriate contemporary evidence before being embedded in present-day operational targets.

The third stage should investigate the drivers of unusually high or unexpectedly low consumption. Monthly fuel use, occupied schedules, conditioned floor area, and equipment operation can distinguish weather-sensitive demand from persistent baseloads more effectively than an annual total. These additional data still require interpretation: a high summer load could reflect a fault, summer instruction, community activity, or another legitimate use. Benchmarking is therefore most productive when it generates specific questions for facilities staff and occupants. It should narrow the investigation without presenting an unexplained residual as a finding of managerial failure.

Investment decisions should combine this performance evidence with building condition and the ability to implement improvements. The U.S. Government Accountability Office (2020) documents widespread reported needs to update or replace school building systems, indicating that facilities priorities extend beyond an energy ranking. A building can have modest EUI while possessing unreliable equipment, and a high-EUI building may be providing essential services with inadequate maintenance resources. Energy screening can help locate opportunities, but condition assessments, project feasibility, and educational requirements should determine which interventions are justified and how they are sequenced.

Fairness across climate groups requires neither ignoring climate nor allowing it to explain away all high use. An unadjusted threshold can burden cold-region schools simply because their historical consumption distribution is higher. Excessively generous normalization, however, can absorb remediable deficiencies into an expected-use model and weaken incentives to improve. A practical benchmark should therefore disclose the raw outcome, the adjustment, and the residual separately. That transparency allows users to see whether a favorable score reflects low absolute consumption, difficult operating conditions, or assumptions in the comparison model.

Fuel-specific reporting should accompany any effort to change heating systems or decarbonize school operations. A transition can lower major-fuel site consumption while changing electricity demand, costs, and emissions in different ways. The present findings demonstrate the importance of those distinctions but do not estimate the effects of electrification or prescribe particular equipment. Such decisions require local technical and economic assessment using appropriate emissions and tariff information. The role of this study is to show why the initial benchmark must retain the information needed for that later assessment rather than hiding the fuel mix inside one total.

Finally, evaluation should continue after a project is completed. A post-project reduction in annual EUI is suggestive, but its interpretation depends on weather, occupancy, floor-area changes, and service continuity. A credible assessment compares consistent boundaries over time, accounts for relevant operating changes, and verifies indoor conditions alongside consumption. This is a stronger evidentiary requirement than the cross-sectional analysis can meet, and it is appropriately assigned to project evaluation rather than retroactively claimed here. The proposed sequence connects national screening to local diagnosis and measured follow-through without confusing those stages.

7. Limitations and research priorities

The principal temporal limitation is that the analysis describes 2012. A stable historical file facilitates transparent reconstruction, but it cannot reveal the effects of later equipment changes, solar installations, electrification, or altered school schedules. Its usefulness is therefore methodological and historical rather than a claim to current national intensity levels. Future research should seek newer, comparably documented evidence with sufficient public climate detail or authorized access to confidential survey data. Until that evidence is available, replacing the reference year in a table heading would not constitute an update to the findings.

The ownership and building definitions also limit coverage. Government-owned elementary/middle and high-school buildings are an identifiable statistical domain, but they do not encompass every arrangement under which publicly funded education occurs. Separate buildings on campuses can also represent quite

different functions, even when they share an educational classification. These boundaries may contribute to the observed distribution and especially to the gap between building-weighted and area-based summaries. A future linkage to school administrative records would be useful only if it preserves metering and physical boundaries and respects the survey's confidentiality protections.

The public climate classification constrains environmental interpretation. The combined dry/hot-humid category cannot reveal separate humidity effects, and the marine sample cannot sustain an independently reported estimate. Broad categories also conceal local weather variation and interactions with building design. These restrictions are not resolved by attaching more specific labels to the same observations. Future evidence should permit analysis of thermal and moisture conditions at appropriate resolution, while retaining enough independent buildings to support reliable comparisons and avoiding attempts to infer protected locations from masked data.

Survey-derived confidence intervals address sampling uncertainty but not every uncertainty in consumption, floorspace, imputation, or the analytical model. The independently reproduced EIA example verifies the calculation, and the influence check reduces concern about a single dominant record. Neither check establishes that measurement errors are absent or independent of climate and school characteristics. The confidence intervals should consequently be read as one component of the evidentiary assessment. More detailed meter validation, sensitivity to processing decisions, and direct comparison with audited building records would strengthen future analyses where those data can be obtained legitimately.

The adjusted model remains exploratory and has limited explanatory power. Its predictors are observed characteristics, not randomly assigned treatments; heated and cooled proportions may mediate climate responses, and unmeasured service differences remain. Exponentiating a log coefficient also does not recover the arithmetic portfolio ratio without further assumptions. The model is therefore reported to clarify conditional associations and their uncertainty, not to deliver a deployable school-rating algorithm. Developing such an algorithm would require predictive validation, examination of calibration across school populations, and explicit decisions about which service differences should be normalized.

The absence of direct indoor-environmental, educational-output, cost, and emissions measures prevents a comprehensive assessment of public value. The analysis can identify energy patterns but cannot establish whether a particular school is adequately ventilated, financially efficient, or a priority for carbon reduction. Those outcomes should be investigated through linked, appropriately governed evidence rather than inferred from EUI. A productive research agenda would combine annual and monthly consumption with verified physical boundaries, use schedules, system characteristics, and selected service indicators, enabling progressively stronger questions while preserving the distinction between description and causal evaluation.

8. Conclusion

The 2012 public-use evidence shows a clear historical gradient in major-fuel site EUI among government-owned school buildings: approximately 67.2 kBtu/ft²/year in very cold/cold climates, 56.8 in mixed-humid climates, and 46.8 in the combined hot-dry/mixed-dry/hot-humid group. The gradient persists after school-subtype standardization, while electricity intensity follows the opposite ordering and the building distributions overlap. These results establish meaningful differences in the observed stock without identifying a pure climatic effect or the savings achievable by individual schools.

The central implication is that climate-sensitive benchmarking should be designed as an evidence-based screening process with an explicit population and energy boundary. Portfolio ratios, building distributions, fuel components, and adjusted associations supply complementary information; none can stand in for all the others. Educational service and indoor conditions remain necessary to interpreting improvement. By preserving these distinctions and providing reproducible calculations with survey-design uncertainty, public benchmarking data can support more credible school-energy decisions while making clear which questions require additional evidence.

References

Borgstein, E. H., Lamberts, R., & Hensen, J. L. M. (2016). Evaluating energy performance in non-domestic buildings: A review. Energy and Buildings, 128, 734-755. https://doi.org/10.1016/j.enbuild.2016.07.018

Chan, W. R., Li, X., Singer, B. C., Pistochini, T., Vernon, D., Outcault, S., Sanguinetti, A., & Modera, M. (2020). Ventilation rates in California classrooms: Why many recent HVAC retrofits are not delivering sufficient ventilation. Building and Environment, 167, Article 106426. https://doi.org/10.1016/j.buildenv.2019.106426

Daly, D., Roth, J., Kokogiannakis, G., McDowell, C., Tibbs, M., & Cooper, P. (2022). Energy consumption in Australian primary schools: Influences and metrics. Energy and Buildings, 277, Article 112549. https://doi.org/10.1016/j.enbuild.2022.112549

Dias Pereira, L., Raimondo, D., Corgnati, S. P., & Gameiro da Silva, M. (2014). Energy consumption in schools - A review paper. Renewable and Sustainable Energy Reviews, 40, 911-922. https://doi.org/10.1016/j.rser.2014.08.010

Fisk, W. J. (2017). The ventilation problem in schools: Literature review. Indoor Air, 27(6), 1039-1051. https://doi.org/10.1111/ina.12403

Haverinen-Shaughnessy, U., & Shaughnessy, R. J. (2015). Effects of classroom ventilation rate and temperature on students' test scores. PLOS ONE, 10(8), Article e0136165. https://doi.org/10.1371/journal.pone.0136165

Hong, S.-M., Paterson, G., Mumovic, D., & Steadman, P. (2014). Improved benchmarking comparability for energy consumption in schools. Building Research & Information, 42(1), 47-61. https://doi.org/10.1080/09613218.2013.814746

Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. Applied Energy, 197, 303-317. https://doi.org/10.1016/j.apenergy.2017.04.005

Ouf, M., Issa, M., & Merkel, P. (2016). Analysis of real-time electricity consumption in Canadian school buildings. Energy and Buildings, 128, 530-539. https://doi.org/10.1016/j.enbuild.2016.07.022

Palmer, K., & Walls, M. (2017). Using information to close the energy efficiency gap: A review of benchmarking and disclosure ordinances. Energy Efficiency, 10(3), 673-691. https://doi.org/10.1007/s12053-016-9480-5

Papadopoulos, S., & Kontokosta, C. E. (2019). Grading buildings on energy performance using city benchmarking data. Applied Energy, 233-234, 244-253. https://doi.org/10.1016/j.apenergy.2018.10.053

Sekki, T., Airaksinen, M., & Saari, A. (2015). Impact of building usage and occupancy on energy consumption in Finnish daycare and school buildings. Energy and Buildings, 105, 247-257. https://doi.org/10.1016/j.enbuild.2015.07.036

U.S. Energy Information Administration. (2016a). 2012 Commercial Buildings Energy Consumption Survey public use microdata (August 2016 revision) [Data set]. https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata

U.S. Energy Information Administration. (2016b, August). User's guide to the 2012 CBECS public use microdata file. https://www.eia.gov/consumption/commercial/data/2012/pdf/user_guide_public_use_aug2016.pdf

U.S. Energy Information Administration. (2023, January). User's guide to the 2018 CBECS public use microdata file. https://www.eia.gov/consumption/commercial/data/2018/pdf/Users%20Guide%20to%20the%202018%20CBECS%20Public%20Use%20Microdata%20File.pdf

U.S. Environmental Protection Agency. (2018, August). ENERGY STAR score for K-12 schools in the United States [Technical reference]. https://www.energystar.gov/sites/default/files/tools/K_12_August_2018_EN_508.pdf

U.S. Environmental Protection Agency. (2020, August). Climate and weather [Technical reference]. https://www.energystar.gov/sites/default/files/tools/Climate_and_Weather_2020_508.pdf

U.S. Government Accountability Office. (2020, June). K-12 education: School districts frequently identified multiple building systems needing updates or replacement (GAO-20-494). https://www.gao.gov/products/gao-20-494