Empirical Benchmarking of Building Performance

Paul Bannister, Exergy Australia
Adam Hinge, Sustainable Energy Partnerships

ABSTRACT

In a variety of voluntary and regulatory initiatives around the globe, including the introduction of the European Building Performance Directive, the question of how to assess the performance of commercial buildings has become a critical issue. There are presently a number of initiatives for the assessment of actual building performance internationally, including in particular US ENERGY STAR® Buildings rating tools and the Australian Building Greenhouse Rating scheme. These schemes seek to assess building energy performance on the basis of actual achieved results, which takes into account not only the theory of how well a design works but also the quality and fidelity of delivery, commissioning, operation and maintenance. It has been shown that in virtually identical buildings, with very similar systems, equipment, and space usage patterns, differences in control, operation and maintenance can generate dramatic impacts on the energy performance of the building.

Given the known dissonances between theoretical performance and actual performance true operational performance-based assessment is essential if policy initiatives are to be assured of delivering actual benefits. Such assessment also has the strong benefit of being somewhat cheaper than design based approaches. In this paper, the various empirical benchmarking schemes around the world and in development are identified and briefly scoped. A summary of the benefits and problems of empirical benchmarking are discussed, and solutions and examples drawn from key existing schemes are presented.

Introduction

In the past decade there has been increased attention toward understanding the energy performance of buildings, with particular interest being paid to the potential to significantly reduce building energy consumption. Given the known dissonances between theoretical performance and actual performance – such as documented in the PROBE studies (e.g. Bordass et al 1996, Bordass 2004) - true operational performance-based assessment is essential if policy initiatives are to be assured of delivering actual benefits. This reflects the reality that actual building performance is a product not just of design but also of design delivery, commissioning, operation and maintenance.

There are several initiatives for the assessment of actual building performance internationally, including in particular US ENERGY STAR® Buildings rating tools and the Australian Building Greenhouse Rating (ABGR) scheme. Furthermore, a number of schemes are under development in Europe in response to the European Building Performance Directive.

In this paper, the benefits, applications and issues associated with the establishment of performance based assessment based on empirical benchmarking are reviewed, drawing on examples from existing schemes.

What Is an Empirical Benchmark?

Empirical benchmarking, as relevant to this paper, is comparison of actual building performance – typically applied to energy – against the broader building market to achieve a view as to the performance of the building.

In creating such a benchmark, one has to be sure of the comparability of the building to the data set. This typically requires that a range of normalization factors are required to bring both back to a common basis for comparison. By this method, differences in climate, building size and hours of operation can typically be eliminated from the comparison.

Significantly, empirical benchmarking approaches characterize performance outcomes, but do not seek to characterize how that performance is achieved.

A number of the key public empirical benchmarking systems are summarized in Table 1. A variety of other benchmarking systems and schemes also exist; Table 1 is not intended to be comprehensive but more representative of the schemes in use around the world.

The Relationship Between Empirical and Design Based Assessments

There are two basic ways in which a building's performance can be assessed. These are:

- **Design:** identifying the presence or otherwise of features that are deemed to be associated with efficiency. This may be done via a checklist and/or via the use of computer simulation, and although it most typically applies to new buildings, may be applied to existing buildings as well. Examples of this type of approach are LEED-NC, BREEAM, ASHRAE 90.1 and other building codes.
- **Empirical Performance:** looking at actual building performance and determining efficiency on that basis. This is typically applied to existing buildings but can also be applied to new construction. Examples of this approach are ENERGY STAR® in the US and the Australian Building Greenhouse Rating scheme.

Design based assessments are a natural response to the problem of assessment prior to construction. However, such approaches carry a number of disadvantages:

- There is no guarantee that the efficiency result will occur. The pivotal study in this area was undertaken by the PROBE team in the 1990s (e.g. Bordass et al 1996, Bordass 2004). In the US, a review of a large data set of commercial buildings found substantial differences between simulated design intent and as-built performance (New Buildings Institute 2003).
- The limitation to those factors which can be readily assessed may mean that other, critical but difficult to characterize or monitor issues are ignored. For example, while design assessment systems may make reference to the need for good commissioning, the ability to assess this stops short of actually determining whether the scope and execution of the commissioning and subsequent rectifications have actually resolved the issues affecting efficiency.

Table 1. Summary of Public Benchmarking Schemes

Scheme	Sector	Description
Australian Building Greenhouse Rating Scheme	Office buildings – land lord services (air-conditioning, common area services, tenant car parks) Office buildings – tenancies (tenant light and power) Office buildings – whole buildings Retail shopping centres (under development) Hospitals (under development)	Five band star rating scheme with 2.5 stars set to population median. Assesses greenhouse gas emissions related to energy use. Scheme established in New South Wales in 1999 and extended nationally across Australia in 2000. Benchmarks adjusted in each state to allow for different local population characteristics. Typical data sample per state approximately 60-200 buildings. References: www.abgr.com.au, Bannister et al 2004.
National Australian Built Environment Rating System	Office buildings – water Homes – greenhouse emissions Homes – water	Five band star rating scheme with 2.5 stars set to population median. Assesses greenhouse gas emissions associated with energy use and water consumption in kL per m^2 for offices and kL per occupant for homes.
Hong Kong Energy Benchmarking Tool	Office buildings – whole building Office building – tenancies Office buildings – landlord services Retail (several types) Hotels Boarding houses Universities Post-secondary colleges Schools with air-conditioning Schools without air-conditioning	Statistically based system that reports position in the market place in terms of percentiles. Assesses energy use. Scheme has been operating for several years during which time it has expanded to cover a wide range of building types. References: EMSD 2006
US ENERGY STAR® for Buildings	Office Buildings – whole buildings Hotels K-12 Schools Medical offices Supermarkets Warehouses Hospitals Dorms	Rates buildings on a 1-100 scale, based on weather normalized primary energy consumption per year per square foot of floor area. Voluntary scheme developed by the US Environmental Protection Agency (EPA), began for office buildings in 1999 and additional building types have been added since. Candidate buildings are compared to similar buildings from a large dataset, the US Commercial Buildings Energy Consumption Survey (CBECS), the most comprehensive measured building energy performance database in the world. Over 18,000 buildings have been benchmarked, representing approximately 300 million square meters, or about 5% of the total US commercial building stock. References: www.energystar.gov Hicks et al 2004.
Cal-Arch	All commercial buildings (non-residential, non-industrial)	Cal-Arch is a distributional benchmarking tool developed by Lawrence Berkeley National Laboratory, a research laboratory affiliated with the US Department of Energy. The Cal-Arch system utilizes a California specific dataset of energy use, and allows users to determine the percent of similar buildings that use more or less energy than their building. A user can compare their building EUI (energy use intensity) to that of similar buildings in the same climate zone. References: CEC Energy Benchmarking White Paper 2005

Design approaches can also be expensive. An ABGR assessment costs typically in the region of $1,500-$2,500 (US$1,000-1,500). Design assessments generally cost several times this figure.

If a design-based approach is extended to the environment of existing buildings, further issues appear:

- The required design information may not be available (for instance, getting detailed information on glazing is difficult without actually sending a pane off for laboratory testing)
- Key factors determining existing building performance may not be practically assessable from design information only – particularly in relation to operation and maintenance.

One of the key factors that arises in the review of design based rating systems is that most have been developed specifically to assess the design independently of operational factors. This approach makes good sense for building codes and other areas where there are distinct limits to what can be included within the scope of an assessment. However, if one is interested in changing the performance-based outcomes of the design, a broader approach is required.

By contrast, empirical performance based approaches have arisen largely out of the existing buildings sector. Such approaches have focused on available data – energy bills, floor area, hours of operation, etc. This is simple data for existing buildings but is problematic for new buildings, for which such data need to be estimated or simulated.

The key benefit of empirical approaches is that by focusing on actual measured building energy performance, one is working with actual performance outcomes. This means that all of the various factors that may have contributed to building performance – design, commissioning, operation, maintenance – are captured in one relatively simply obtained figure. In an environment where there is a need to manage actual building performance for cost or environmental reasons, this is hard to substitute with design based assessments, which provide only what we *think* will work, as opposed to what *does* work. The key disadvantage of an empirical performance approach is that the assessment provides no information on what is or is not working, unless a further and deeper analysis is performed. This means that a poor empirical performance has to be followed up by additional investigations before causes can be ascribed and resolved.

The value of empirical benchmarking is becoming more widely accepted as a result of the voluntary initiatives described in this paper. In the State of California, an Executive Order was issued in 2004 requiring a plan "to accomplish benchmarking of all commercial and public buildings in California, including benchmarking at the time of sale, as well as a system by which benchmarking ratings can be disclosed to tenants, buyers and lenders to advise them in making decisions" (CEC Benchmarking White Paper 2005). In Australia, studies are underway to consider the possibility of a compulsory building labeling system, with one of the options being performance based. Several Australian State Governments are already requiring minimum ABGR performance levels for buildings that they lease.

It would be fair to say that there is some tension between the design assessment and performance benchmarking approaches, particularly as designers do not wish to be held responsible for in-practice performance that includes many factors outside their control. However, with the proliferation of "Green" or "High Performance" building design around the world, serious work is required and in some cases underway to rationalize the actual performance of many of these buildings. Much more work is needed to understand the reasons for disappointing performance of "well-designed" buildings.

Irrespective, it is important to emphasize that design and empirical performance assessment should not be viewed as being in conflict. Design assessment provides an important role in systematizing our understanding of good design, while performance assessment provides feedback on what works. The ideal operation of these two systems is for the knowledge gained through the use of performance assessment to be passed back and incorporated into the design assessment systems. In this way, the differences between these two systems will be minimized via the delivery of better buildings.

Value of Empirical Benchmarking Schemes in the Market Place

Market Transformation

In an ideal world, we would all make choices based on perfect information, and one of the factors we might choose to use in sourcing/leasing a building could be efficiency. Benchmarking systems have a potentially critical role to play in this area, by providing information that enables a common language of comparison to be used throughout the market. However, if market transformation is to be successful, it is important to be able to break the rating into components that can be managed by individual market participants and then used reasonably to broker their relation with other market participants.

The principal benchmarking systems in use to date differ in whether they benchmark "whole building performance", looking at all of the energy uses in the given building, or break the energy use benchmark into the portions controlled by different decision makers.

This latter approach is used in the Australian Building Greenhouse Rating scheme and the Honk Kong Benchmarking Tools (EMSD 2006), which provide two separate ratings as listed in Table 1. The base building rating has been set to reflect a relatively frequent metering configuration whereby the "landlord services" – air-conditioning, lifts, common area lighting and tenant car parks - are captured by one electricity meter and sometimes grossed at a fixed rate into the rent.

With this information, a building owner can then advertise the performance of their building in a manner which is largely independent of their tenants. Indeed, simulation studies have shown that in the Sydney climate at least, a range of tenancy energy consumption of a factor of more than four has only a 0.6 star effect (approximately 10% of scale) on the base building rating.

In Australia, the use of ABGR has been accepted widely and its role in transforming the market is widely accepted. The central impact has been around the brokering of the landlord-tenant relationship, with government tenants requesting 3.5 star and higher base buildings and developers endeavouring to achieve 4 stars and higher in new construction.

The key power of this approach is that efficiency becomes linked to a far stronger factor – in this case, lettability (leasibility) – with a significant leveraging effect on the market's level of interest and activity in relation to efficiency.

ENERGY STAR® by contrast does not provide a base building tenant differentiation, in part due to less homogeneity of lease structures around the US. Nonetheless, with increasing awareness of the ENERGY STAR® rating system, along with a growing sense of urgency toward energy savings in some parts of the country, the transformation of the market toward benchmarking is making great progress. As noted earlier, following the results of voluntary efforts, the State of California has recently mandated that all public and commercial buildings in the State will be benchmarked within a few years. As this initiative continues into implementation, other states in the US are likely to develop similar requirements.

Technical Information

The role of empirical benchmarking in providing technical information is equally critical but more complicated.

At a basic level, empirical benchmarks provide only limited resolution of differentiation between buildings – effectively "good", "bad" or "average" – and thus provide a crude indication of the technical status of a building.

Furthermore, as noted earlier in this paper, the available empirical benchmarking systems do not correct for levels of service or differing amenities. Thus a building that provides only limited space temperature control, or perhaps provides fewer other services than the general comparison data set will perform better than would otherwise be expected given the technology, operation and maintenance of the building.

However, given that there is demand for empirical benchmarks even in sectors where there is no market mechanism in operation (e.g. hospitals, schools), it is clear that an empirical performance assessment can provide useful information. Underlying this is the reality that the range of actual building performance driven by manageable efficiency related issues is generally far greater than can be ascribed to all but the most obvious (and therefore easily identified) differences in service.

Experience with the Australian Building Greenhouse Rating scheme has identified a wide range of technical applications to the various ratings. From a technical perspective it has been repeatably shown that poor ratings indicate the existence of genuine efficiency issues, even where energy audits have previously failed to detect these.

However, the success of this type approach is strongly affected by the relevance and accuracy of normalization factors, and of course the quality of the underlying data. A poorly formulated benchmark, with poor normalizations or dubious underlying data, has a high risk of providing misleading or incorrect information. This of course has implications for both the market and technical uses of the rating.

What arises from this from a technical perspective is that there is a strong need for the "meaning" of different positions relative to a benchmark or a rating to be demonstrated, recorded and made public. It is only in this manner that a fully qualified understanding of the technical meaning of the rating can be developed. Clearly for the rating to be successful it must be clearly demonstrable that a good rating is genuinely efficient and not just lucky.

This has been the partial experience under ABGR, where there is a reasonable amount of knowledge built up on the rating and the "meaning" of the various ratings, and it has been shown that the rating is a good indicator of actual efficiency. However, the degree of knowledge at this level in the broader market place is limited and it has been left very much up to individuals to determine this knowledge for themselves.

New Buildings

The role of empirical benchmarking in new buildings is more complex again, as the obviously the building is not yet operating.

Irrespective of this, it is very likely that, given a successful benchmarking system for existing buildings, building project clients will start demanding that these benchmarks are met for new buildings – after all, a new building is, upon handover, just an existing building.

This creates a significant challenge for the design and construction industry, as there are no well defined processes for the assessment of actual performance during the design process. The most obvious – computer simulation – is not generally geared towards absolute energy use prediction and there are significant issues in the reconciliation of this against actual performance.

Furthermore, in a situation where the "cutting edge" may be at levels of performance that have not yet been achieved in any building, the challenge of how to meet such a target is clearly problematic. This situation is made worse when tenancy organizations seek high performance levels but then also specify energy-wasting requirements within the balance of the brief that are directly incompatible with the achievement of the goal.

These problems notwithstanding, it is reasonable for the market to expect delivery of buildings that actually work to a defined performance level as long as the parameters of performance are reasonably within the control of the design and construction team. Thus for example a "landlord services" style benchmark is reasonable as this can largely be controlled by the design/construct team, while a tenancy or whole building benchmark is not, as the tenants may significantly affect this. Compliance with a tenancy or whole building benchmark performance places significant requirements upon the tenant as well, as tenancy energy consumption is as much about internal office system and operation as it is about lighting design.

To manage the various issues, a number of market information issues need to be addressed:

1. Provision of advice to tenancy organizations on what level of performance to reasonably expect for new buildings.
2. Provision of advice to tenancy organizations on how to specify target performance levels and what other brief items may be incompatible with this.
3. Provision of advice to designers and construction organizations on what the ratings mean, and any existing precedents for their achievement
4. Development of processes and information to assist all participants in the delivery of performance targeted projects, drawing on the best knowledge available, and updated as experience increases.

These items have the potential to make a significant difference to the market response – and the possible levels of inter-party conflict – to the use of performance based requirements for new buildings.

General Characteristics of Empirical Benchmarks

The following sections review the key technical considerations required when establishing a benchmarking system.

The Statistics of Building Populations

The nature of most commercial building types is that a range of core, or landlord, services such as HVAC, common area and exterior lights, elevators, etc., are provided in all buildings, with the result that the general distribution type is normal but with a high energy use tail. This reflects the reality that the provision of those core services takes a finite amount of energy, and there are technological limitations as to how efficiently the broad building market is capable of providing those services. As the average energy use can be heavily affected by a relatively small number of high energy samples, it is generally preferable to use the median to characterize the midpoint of the distribution.

A further feature of any real data set is the presence of erroneous data. These data are particularly prevalent at the extremes of the population, so that any methodology based on real data needs to recognize the potentially flawed nature of such data. In particular, the data at the extremes of the performance spectrum should be treated as untrustworthy, as experience shows that this is often the case. This however makes the definition of "high efficiency" relative to the market more challenging.

A good review of statistical issues related to benchmarking is contained in the recent report "Review of California and National Benchmarking Methods" (Matson & Piette 2005).

Normalization Factors

In order to make a meaningful comparison of a building against the broader population, it is essential to define a common basis of comparison that allows the efficiency-related issues (such as poor plant, maintenance or operation) to be differentiated from primary factors (such as building size) or operational factors (such as hours of servicing). This requires the introduction of normalization factors that are used to correct a building back to a reference case of "normal" operation. Thus an office building can be characterized in terms of an energy per m^2 (or per square foot) figure that has been corrected to a standard climate and operating hours. This then permits that building to be compared reasonably against other buildings.

When normalizing for the issue of size, it is important to normalize back to a relevant variable that is closely linked with the productive function of the building. Thus for instance, normalization to net lettable area in an office provides a better recognition of the productive variable for the office, whereas gross conditioned space may include significant areas of the building that are not lettable and thus not productive. Similarly, hotels and hospitals are arguably better characterized in terms of beds rather than floor area, as it is the beds in both cases that are associated with production, rather than the floor space.

Creation of Normalization Factors

There are two basic methodologies for producing normalization factors:

1. Empirical derivation, via the identification of trends in the building population correlated to individual normalization factors; and
2. Theoretical derivation, via the identification of trends predicted from simulation modeling or other such theoretical approach.

Each of these approaches has advantages and disadvantages, and the selection of the most appropriate methodology needs to be pragmatic.

An empirically derived normalization approach has the key benefit that the results are based on actual performance data. However, the validity of such an approach is dependent upon whether the data set has sufficient data with independent variation across each of the variables. Thus for instance, it is difficult to get a large sample of data across the full range of operating hours for office buildings, as the number of office buildings with more than 70 hours per week operation is small by comparison with the balance of the population. Similarly, there has been little evidence of empirical correlation between climate and building performance in most of the studies that have been performed.

In an interesting aside, although energy use typically fails to correlate with climate to any significant extent, recent work in Australia (Bannister et.al., 2005) has identified a significant correlation between the climate and water consumption. Indeed, water consumption in buildings between Melbourne (a mild temperate climate) and Brisbane (a sub-tropical climate) doubles. Furthermore, extrapolation to UK climates via cooling degree days brings reasonable comparability to UK benchmarks. This climate correlation is believed to be the result of increased cooling tower water consumption, creating the interesting possibility that it may be realistic to derive a climate correction for cooling energy on the basis of water consumption.

A further point of note on the issue of climate correction is that studies in Australia have repeatably shown that dry-bulb cooling degree days are inappropriate as a basis for climate comparison. Cooling loads are far more strongly correlated to wet-bulb temperatures, with the result that wet-bulb cooling degree day data is a better basis for normalization. However, such data are not routinely available and there are issues with the consistency of calculation.

In general, however, the reliance on statistics alone to generate normalizations is flawed in that the normalizations identify the behaviour of the overall population rather than that of an individual building. Where correlations are strong these two items may be coincident, in which case the empirical normalization is the best methodology; however, for weaker correlations the other changes in the underlying data set may be more important in determining the correlation than the changes in the individual buildings. In this instance the derived normalization typically under-corrects.

Theoretically derived normalizations have the advantage of being derived on an individual basis but the disadvantage of being theoretical and thus potentially wrong. In particular, it is generally necessary to make a specific decision as to whether a theoretical normalization is going to correct for “best practice” or “normal” behaviour. This is not only a value judgment but is also a potential source of error as neither of these is necessarily well defined.

Benchmarking or Rating?

The traditional benchmark approach sets a single figure that is considered, typically either as “average” or “best-practice” performance. This process is utilized in the US ENERGY STAR® scheme, which sets a single “pass” mark on the basis of the 75th percentile of the building population. This has the advantage of providing a clear signal of good practice, but has the disadvantage of discouraging or alienating those parts of the population for whom the achievement of the benchmark appears remote. This is one rationale for why ENERGY STAR®

has recently begun to recognize a ten percent improvement, not just attainment of the 75th percentile.

The alterative approach is to establish a rating, which provides most buildings with a place on a scale. This approach is adopted in the Australian Building Greenhouse Rating scheme, which places buildings on a nine point scale (one to five stars, with half stars), and in Hong Kong, which places buildings in the relevant decile of the population. This approach has the advantage of being more inclusive, such that poorly rating buildings can still use the scale to identify and promote performance without necessarily being best practice. However, it requires some care in informing the market place as to the meaning of the ratings.

Is Good Performance the Same as Efficiency?

A key issue for any performance benchmark is the question of what the results mean. All benchmarking systems reviewed in this paper assess performance solely in terms of utility consumption without necessarily assessing the level of service. This creates the possibility that a building achieves well by failing to provide service. The various schemes available display a range of responses to this:

- **Categorization.** This approach is used in the UK Econ 19 (Energy Efficiency Best Practice Programme 2000), which provides different benchmarks for different levels of building "quality", as indexed by the level of amenities in very broad terms.
- **Normalization.** An alternative approach would be to physically measure or otherwise characterize service on a scale and normalize for it on a mathematically continuous basis, as one might for hours of occupancy. However, the authors are not aware of any schemes that provide any form of continuous correction for service. This is at least in part because there is no necessary agreement internationally on the boundary between necessary service – which would merit normalization - and excessive service – which would not.
- **Limitation.** The approach used by the US ENERGY STAR® scheme is to require a minimum level of service. This provides an assurance that a building does not achieve a high energy rating on the basis of service deprivation. However, there may still be disagreement about what constitutes a reasonable minimum level of service.
- **Avoidance.** The final approach is to clearly identify that all the rating does is rate energy, and let the relevant markets decide on this issues of servicing. This approach is used by the Australian Building Greenhouse Rating scheme and the Hong Kong Benchmarking Tools (EMSD 2006). This methodology is most readily applicable where there are relatively uniform expectations in the market place as to the level of service expected. This is the case in Australia, where basically all commercial buildings are air-conditioned, but would be more problematic in the UK, where the presence or otherwise of air-conditioning is a significant differentiator. Nonetheless, the avoidance approach works well where the function of the rating is to encourage people to use less resources.

Overall, it would appear that the resolution of this issue is dependent on the nature of the specific rating and market under consideration.

Future Directions

Activity in the empirical benchmarking field lacks any form of international cooperation and is characterized by an inefficient process of learning and repeating each other's lessons and mistakes rather than learning from these. There is a need to elevate activity in this field to achieve the following key outcomes:

- Greater data and lesson sharing internationally;
- An improved understanding of the benefits, challenges and limitations of the available processes;
- An improved understanding of the meaning of benchmark results in different climates, countries and subsectors;
- Collaboration with and feedback to design assessment systems;
- Improved industry acceptance and use.

This paper has been prepared with the intention of promoting interest in the creation of activities to meet these needs. It is suggested that the major international professional organizations (e.g ASHRAE and/or CIBSE) would be ideal forums for such activity.

Conclusions

Empirical benchmarking provides a strong basis for the assessment of actual building performance and is essential if actual energy use and greenhouse emissions are to be reduced internationally.

The creation of good quality benchmarks requires good quality data plus the use of normalization factors to separate operational issues from efficiency issues. Benchmarks may take the form of an individual "pass/fail" or a banded system ranging from poor practice to best practice. In either case, it is important that knowledge is built up and disseminated as to what performance at a given benchmark level actually means, so that industry can act upon this information.

Benchmarks have the potential to be used in market transformation, by linking performance to market expectations when leasing, as a technical tool, and in the prescription of new buildings.

It has been proposed that there needs to be greater coordination and promotion of empirical benchmarking activities internationally to ensure that lessons and information are shared to the benefit of all.

References

Bannister, P., and Salmon, S. 2004. "Australian Building Greenhouse Rating Scheme: Rating Operational Energy Efficiency in Office Buildings" in Proceedings of the 2004 Improving Energy Efficiency of Commercial Buildings Conference, April 21-22, 2004. Frankfurt, Germany.

Bannister, P., M. Munzinger and C. Bloomfield. 2005. Water Benchmarks for Offices and Public Buildings. Exergy Australia. 19 September 2005, Belconnen Australia.

Bordass, W, Leaman, A. and Field, J, 1996. "PROBE 7: Homeowners Friendly Society", CIBSE Building Services Journal 39-43 (October 1996).

Bordass, W 2004. "Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap" " in Proceedings of the 2004 Improving Energy Efficiency of Commercial Buildings Conference, April 21-22, 2004. Frankfurt, Germany.

CEC (California Energy Commission) Benchmarking White Paper 2005. Executive Order S 20-04: Benchmarking White Paper. CEC-400-2005-010. March 2005. (also, Benchmarking System for California Commercial Buildings: Plan, Timetable, and Recommendations", CEC-400-2005-051-CMF. September 2005.) Both available at www.energy.ca.gov/greenbuilding

Electrical and Mechanical Services Department (EMSD), Hong Kong Government 2006. See www.emsd.gov.hk/emsd/eng/pee/benchmarkingtool.shtml

Energy Efficiency Best Practice Program 2000. Energy Consumption Guide 19 Energy Use in Offices, UK Energy Efficiency Best Practice Program 2000.

Hicks, T.W. & B. Von Neida 2004. "US National Energy Performance Rating System and ENERGY STAR Building Certification Program" in Proceedings of the 2004 Improving Energy Efficiency of Commercial Buildings Conference, April 21-22, 2004. Frankfurt, Germany.

Hinge, A., J.Rutherford, D.Abrey, J. DaSilva, E. Titus, and E. Smyth. 2002. "Back to School on Energy Benchmarking." In Proceedings of the 2002 ACEEE Summer Study. Monterey, Calif.

Matson & Piette 2005. Review of National and California Benchmarking Methods, Lawrence Berkeley National Laboratory, LBNL Report No. 57364, April 2005.

New Buildings Institute 2003. "Is What They Want What They Get? Examining Field Evidence for Links between Design Intent and As-built Energy Performance of Commercial Buildings". Prepared by the New Buildings Institute, Inc, for the US Environmental Protection Agency. January 2003.