CLOSING THE LOOP
Post-Occupancy Evaluation: The Next Steps
29 April - 2 May 2004
Cumberland Lodge, Windsor, UK

Streamlining Survey Techniques

Adrian Leaman and Bill Bordass

ABSTRACT

Post-occupancy evaluation and building performance analysis were coined in the 1970s as portmanteau terms for clusters of survey techniques applied to the study of buildings in use. Despite development lead times of nearly forty years, monitoring post building performance is still not a routine part of design practice. This is often said to be the reluctance of design professionals to embrace the potentially risky conclusions that studies may bring. However, there are more deep-rooted reasons which are rarely given any attention. These are concerned with the survey methods themselves: their cost, manageability, applicability, replicability and usefulness. This paper deals with how we have tried to overcome some of these methodological hurdles and, in so doing, streamline survey methods so that they are more cost-effective and easier to apply, but still comprehensive in their question-answering abilities, so that their take-up by practice is potentially improved.

KEYWORDS
Building performance evaluation, post-occupancy evaluation, surveys, techniques, methods, data analysis, software tools.

Introduction

WHAT'S THE PROBLEM?
Despite nearly forty years of development, building performance assessment is still not routinely applied to mainstream building design practice. One of the reasons, which has not been given the attention it merits, is that the methods used for assessment can be unwieldy, costly to implement and hard to manage across more than one building study. In brief the problem is that techniques:

- are often created for single-building studies, or single-project, multi-building studies, with little thought for further application later, especially with respect to benchmarking, and use across multi-building types;
- have to consider a large number of variables, with the significance of any of them often not fully clear at the outset, so it is hard to know what to leave out;
- do not take enough account of similar techniques already created for similar study purposes, with a tendency to disregard serviceable methods which are 'not invented here';
- try to achieve too much with the resources available, but also
- grossly underestimate how much of those resources have to be used for back-office tasks like scaling, categorisation and data typing, statistical analysis, software management, and quality control of data inputs and outputs, including range and error-checking.

Added to this are:

- over-optimism about the analytical and presentational capabilities of off-the-shelf software;
- academics tending to look for theories while clients and building designers often need more rough-and-ready guidance (e.g. what should I do in this situation? What should I pay attention to? What should I ask? What should I avoid? How well did that idea work? Can I improve it next time?);
- disagreements over where best to use physical measurements;
- unrealistic expectations about questionnaire design and the internet as a data-gathering tool.

Building performance analysis is a challenging area. There are many pitfalls connected with seemingly mundane tasks of gathering multivariate data, analysing it meaningfully and efficiently, organising it for further use later, maintaining it over time, keeping it error-free, reporting on it effectively, and distributing the results in ways which are attractive and meaningful to specialists and lay people alike.

The upshot can be:

- studies which run excessively over time and budget;
- a tendency, rather like building themselves, for every study to be treated as a prototype;
- emphasis on individual building studies rather than on benchmarking and strategic conclusions across a larger sample of buildings;
- difficulties resolving results produced by different survey teams and protocols;
- 'scale' problems in the use of techniques when moving from single-building studies to studies with larger numbers of buildings of different sizes and functional types;
- over-reliance on standard software and graphical analysis products, with low standards of data presentation;
- a tendency to over-simplify, thereby losing needed detail, or the reverse, a tendency to …

- create data mountains, with excessive amounts of detailed information which is hard or wasteful to analyse.

Streamlining survey methods

OUR APPROACH

This paper deals with our experiences of grappling with some of these problems. It draws examples from techniques with which we have closest familiarity. Our main aim has always been to make survey methods capable of widespread take-up by design practices, building occupiers and researchers so that performance monitoring becomes both a routine part of the design process and a mainstream part of the educational experience of designers. We have tried to make our survey methods more streamlined so that they:

- are relatively less expensive to commission;
- provide unbiased results quickly;
- are useful across a wider range of interest groups, (i.e. studies which are not necessarily restricted to single-issue topics or overly narrow professional audiences);
- cope with the exigencies of the real-world, making allowances for the everyday requirements of building managers and occupants when the studies are carried out;
- are statistically robust, providing believable results that meet stringent statistical and quality control criteria, but are not bogged down in jargon and unnecessary detail;
- are useful in 'mix-and-match' clusters, so that the techniques can be put together in different combinations or separately as stand-alone studies;
- are compatible with other similar studies;
- are less off-putting to carry out for those who may wish to embrace them but may be frightened off by technicalities.

RELATIVELY LESS EXPENSIVE

Cost is the greatest obstacle to take-up. In order to make feedback routine [1], thereby moving moving beyond research and development into everyday design and management practice, costs need to be reduced substantially.

For example, at Building Use Studies (BUS), and without allowing for inflation, the basic cost of a standard occupant questionnaire survey is now one-fifth of what it was ten years ago. This has been achieved by:

- Radically reducing the quantity of information collected, on the basis of 'need to know' not 'nice to have'. The standard occupant survey now has 20 background questions and 45 building performance questions. [2]
- Modularising questionnaires so that sections can be added and subtracted easily.[3]
- Creating a single database for the analysis of all questionnaire variants for every survey undertaken. The current 2004 version of the BUS individual building database has 3,500 database fields. The number of records in the database depends on the size of the sample of occupants for the particular study building. [4] Although the database is relatively large and carries a lot of redundancy because not all of the analytical power is needed for each study, it is much easier to manage across all studies.

- Developing in-house analysis software [5] to carry out all mainstream data analysis and presentation routines automatically [6].
- Giving more emphasis to the production of reliable, attractive and easy-to-understand results and less to report writing.
- Using a licensing system, so that users may apply the techniques for themselves (under licence) and then have them analysed by BUS using the software described above.
- Giving relatively more effort to the management of the benchmarking database. The current 2004 BUS buildings (benchmarking) database has 170 records (buildings, of which 50 are used for the latest set of benchmarks) by 256 fields.
- Placing more emphasis on routine metadata management tasks such as consistent variable naming between surveys, data typing and scale definitions. The BUS system includes a variable and scale name database which has 240 records (variable names). This services the 10 survey variants to ensure that all variable names, nomenclature, scale 'directions' and scale descriptions are absolutely consistent across all surveys. Lack of clarity with definitions across individual building surveys, and confusion between individual building variables and benchmark variables can be a major source of inefficiency. This factor alone accounts for many surveys never proceeding beyond prototypes.

All of the above contribute to reductions in labour costs, both for individual studies and for the support of the benchmarking system. A cost which is harder to avoid is the main analysis report. BUS tries to get round this by supplying licensees with all the data analysis (survey results, benchmarks, comments, graphics and sub-type analyses if any) presented in four appendices, thereby separating the bulk of the data from the main report. Clients may then commission BUS to write the main report, or write their own on the basis of the supplied analysis data. Postgraduate student licensees (who receive the service free) always opt just for the appendices. Commercial clients are more likely to commission the report. Separating the data and graphics components from the main report helps to pare costs.

PROVIDE RESULTS QUICKLY

Keeping costs down is vital to take-up, but so also is getting results out quickly without compromising their quality. As soon as a survey has been carried out, clients want to have the results, so the sooner the better in most cases.

We try to match the timescales of the techniques with the needs of the commissioning client (including disruption and access considerations) rather than making them fit in to longer research and development timescales, which may be months or even years. For example, all the methods used in the Probe studies could be implemented relatively quickly:

- The energy and technical module [7], including a detailed technical report, can be concluded in a elapsed time of no more than ten days after the occupier has completed their pre-visit questionnaire [8], including a detailed technical report. With annual energy use data already available, the spreadsheet-based results can produce a first cut in less than a day and detailed results in typically three days. The major drawback with collecting energy data, however, is getting the energy consumption information out of the occupier’s organisation and its energy suppliers in the first place. In some extreme cases, this can take 2-3 months.
- The occupant module takes a day or less to set up, a day to carry out and less than a day to produce substantial analysis results. A report would add two more days, so the total

turnaround time is five days or less.

- The air tightness module [8], used in the latter stages of Probe, takes about 3 hours for blocking off holes and ducts, 1-2 hours for fixing the monitoring equipment to the building, and about 2 hours for the tests themselves. If any further investigation is required after the first measurement (for example, to investigate the reasons why the building did not pass the test) then this might take a further 3 hours. The test usually takes less than a day all told.

Speed of data capture and reporting must not compromise the quality of the results. One of the main ingredients for Probe's relative success was the ability to deliver robust results on all the 20 buildings attempted in the project and fit these into publication schedules. The techniques allowed us to collect data of publishable quality, even if things did not go particularly smoothly.

This confirmed to us the importance of using tried-and-tested methods, rather than trying out too many new things. In Probe, partly for cost and scheduling reasons, we deliberately did not have survey modules which sought to obtain primary data on e.g. space utilisation, aesthetics or costs in use (a particularly difficult area in which to collect high-quality data quickly and efficiently). We found that secondary information generated by the two main survey modules (energy/technical and occupants) supplemented by specialist interviews where necessary, gave us most of the information we needed on these topics; and well beyond. For example:

- the classified list of written comments in the occupant survey identifies issues not specifically covered by the questions themselves; and
- when doing an energy survey one also reviews the quality of construction, installation, commissioning, record information, operation, management and maintenance, and has a chance to talk to occupants, facilities and maintenance staff whilst walking round, seeking information, checking plant and taking spot readings of light, temperature, air movement and so on.

USEFUL ACROSS A WIDER RANGE OF INTEREST GROUPS

One of the key features of tried-and-tested methods is that they have already been simplified. Buildings are complex multivariate systems with a host of interacting physical and human variables, each with a different context, a further complication. It is often hard for researchers to know which variables to include and where to draw the boundaries, especially if they have to cope with the advice of a multi-disciplinary steering group who may introduce additional requirements, sometimes conflicting.

The tendency is for studies to add complexity, rather than pare it down. This adds to the cost of analysis and reduces the likelihood of the study being completed on time and to budget. The key to successful evaluation is knowing what to leave out.

Probe, for example, was primarily built around the energy/technical and occupant modules, both of which had already been in development for ten years. Probe also used a new pre-visit questionnaire and also called on, at various stages, air tightness tests, a water consumption monitoring method [10], physical measurement of energy consumption where metering was absent, a journey-to-work questionnaire [11] and various on-the-spot measures of light, air infiltration, temperature, humidity, power and noise.

This gave us a rich enough picture of performance to be of use to services engineers (the main audience for Probe in Building Services Journal), clients, architects and facility managers. Subsequently, there had been substantial call from universities for Probe studies as teaching materials. If we had tried to meet all these audiences in the first instance, the requirements would probably have swamped us. We stuck to what we knew could be done in the time available, working within stringent publication deadlines.

COPE WITH THE REAL-WORLD

Our approach to performance studies is as pragmatic as possible. We want to get into buildings with little disruption to the occupants, get the data, leave making no traces, then produce believable, useful and interesting results quickly. The results themselves must not be too threatening or make people look silly. The point is to allow simple questions to be honestly answered: How good is this building compared with others like it? If the building works very well, what are the ingredients for success? If things are going wrong, what are they are, can they be fixed, and how can they be prevented in future? We use a maxim adapted from the late Bill Allen: the results of the research should never be more than one step away from a design decision (we now say design or management decision, as the management aspects have become much more important as e.g. technical complexity and intensity of use have increased).

The focus is therefore decision-making, not necessarily hypothesis testing, or theory building. This is one of the reasons why building performance assessment often has an uncomfortable relationship with the academic world: it is not only inter-disciplinary, but strongly rooted in real-world situations and decision support systems, with an emphasis on risk and consequences rather than theory and cause-and-effect. Buildings in use cannot be treated as if they are experimental laboratories with sets of independent variables that can be interfered with at will in order to test outcomes. The effects on, and expectations of, occupants must always be considered. Any survey of occupants which asks about comfort, health, satisfaction and productivity, inevitably raises the respondents' expectations that something will be done about any of the problems they may highlight in their responses. Thus performance monitoring must be linked to responsible and pro-active facilities management in the first instance, and design decision-making which is committed to continuous learning and improvement in the longer term.

STATISTICALLY ROBUST

If techniques are rooted in the real-world, they must also be robust; that is, they must still work reasonably well even if some of the assumptions on which they are based are mildly violated (for example, you may discover on survey day that management will not given you access to the whole building, only to a typical floor). It may not be possible for cost or access reasons to achieve the levels of sampling that might be desirable. We normally try for the largest sample that is reasonably available (especially in buildings with less than 100 permanent occupants). In larger buildings, we try for a minimum of 125 responses. The mean sample size across all the buildings we have studied in the BUS occupant survey is 122. Response rates are almost always better than 80 per cent of those sampled (where BUS administers the survey on site we normally achieve 90 per cent plus). We normally stick to paper questionnaires, personally handed out to and collected, in the building because these reach everyone in the building (including staff who do not have computers) and produce much higher response rates than internet-based questionnaires. High response rates help overcome the possibility of bias, that is, reaching the wrong conclusions because the sample is not representative. With these criteria higher response rates are more important than large samples.

In a different way, this also applies in energy monitoring where the techniques we advocate have:

- graduated response, that is, they can cope with different levels and qualities of input data. For example, energy use and equipment may only be partially available, or there may be uncertainty about the veracity of the data; and there is no point collecting details which you do not really need (we do it on a "need to have" basis instead); and
- in-built error-finding capability, where they can detect discrepancies and/or mistakes in the input data and make them clear in the outputs.

The TM22 energy reporting method used in Probe has a 'double-entry' system to help resolve discrepancies. Metered supply data is calculated separately from the demand data and the balance is reviewed throughout the calculation process. The BUS occupant survey method has separate software for range-checking and variable name checking the spreadsheet data input files. Log files of anomalies systematically list any suspicious input data. This approach is more systematic than using menu-driven spreadsheets, which can be more error prone.

Simple mistakes can lead to gross errors. For example, during a study carried out in 1992 by Building Use Studies, a client asked to change some of the questionnaire scales round (e.g. 'high' became 'low', and 'low' became 'high'). The inexperienced researcher agreed to do this to mollify a persistent client, although this was strictly against normal practice. The variable names were not changed. The analysis was done by someone else who was not aware that some scales had been changed. This meant that some, but not all, of the results became exactly the opposite of what they should have been! The error was subsequently noticed and corrected, but it meant that the database files had to include two versions of each of the calculation fields for all the variables in question. Although seemingly innocuous, this type of alteration can have serious consequences for quality assurance later.

Similarly, the TM22 method has detected errors in highly detailed energy monitoring, e.g.

- heat and electricity meters not monitoring the circuits they were supposed to;
- incorrect meter factors applied;
- gross estimation errors of the energy use of parts of the building not monitored in detail.

It can therefore complement even highly detailed investigations.

MIX-AND-MATCH

Building performance evaluation is rarely the same from one study building to the next. The techniques themselves evolve, requirements may change between evaluation projects with special topics of interest included only for a particular building, and the emphasis given to functions may alter (e.g. one study may require a module for space utilisation, another for energy). The easier it is for researchers to use different techniques in various combinations, the better. In Probe, for example, it was only in the third series of studies that the methods stabilised and settled down, and the production system became more routine. Although it may be desirable for performance assessments to cover as many performance features as possible, it is rarely practicable; or necessary – information overload comes quite early, with more issues arising than anyone can cope with. As the Feedback Portfolio [12] shows, no one technique covers all stages of briefing, design and use, so they must be used in combination to some extent; a conclusion also reached by the Federal Facilities Council in North America [13]

COMPATIBILITY

Ideally, techniques should also be compatible across studies so that comparisons may be more easily made. It makes little sense for researchers to use their own versions of question clusters on e.g. lighting, noise, ventilation and thermal comfort, when most of these are similar from one study to the next – though some surveys may of course require more or less detail, it can be helpful for a set of core questions to be identical. This applies especially when two complementary databases need to be brought together. It is best if there is at least one data field in common. However, as we have shown previously, tiny changes in formats destroy compatibility, even between two similar studies in the same organisation! In theory, open source arrangements based on licensing help to overcome this. In future, researchers and practitioners will make more use of licensing.

Developments in web-based mark-up languages like XML [14] will also have their place. The potential for XML in this area is quite significant. XML is a stricter version of HTML (the main browser display language used on the internet). XML allows computer programs to extract machine-readable data from files. Once a custom markup language has been defined - and a standard language for building performance assessment is not unthinkable - computer programs then can automatically extract tagged data. For data managers, this sounds like a dream come true because it introduces a higher level of standardisation which greatly improves definitional clarity and data capture, just the things that usually reduce survey analysis efficiency. However, as experience with rogue variants of computer languages like HTML, Java and Linux show, even the best and most vigorously defended open-source arrangements are subject to constant undermining from greedy commercial interests.

Building performance assessment has been held back by the absence of standard formats, analysis protocols and definitions for even the simplest terminology. For instance, there are many ways of measuring the different types of space in a building and these can even vary between different professions in the same country (as in the UK)! Energy assessment is another case in point: although more progress has been made in this area. Thirty years after the first oil crisis it is surprising that there are still many misunderstandings and discrepancies over basic terminology and concepts. Just recently, the Europrosper project [15] has developed prototype software for preparing hypothetical energy certificates for offices largely automatically, and including recommended energy-saving measures. This work is still in the development stage, but its approach demonstrates that it is possible to overcome definitional variability by using a carefully-defined common analytical core.

An early version of he strategic framework for Europrosper approach is shown in Figure 1 [14]. It is not necessary to explain the details here, but notice how:

- the various stages of input, throughput/calculation and output are carefully layered (see the rows of Figure 1);

- functional types (e.g. buildings in use types and empty building types) are explicitly defined;
- national reference data are carefully separated from the analysis core;
- production of the output - the energy certificate - is left to the last possible moment, so that calculation phases do not intrude at an earlier stage.

This type of modularisation is an important element of successful technique development.

Figure1: Strategic framework for Europrosper energy certificates

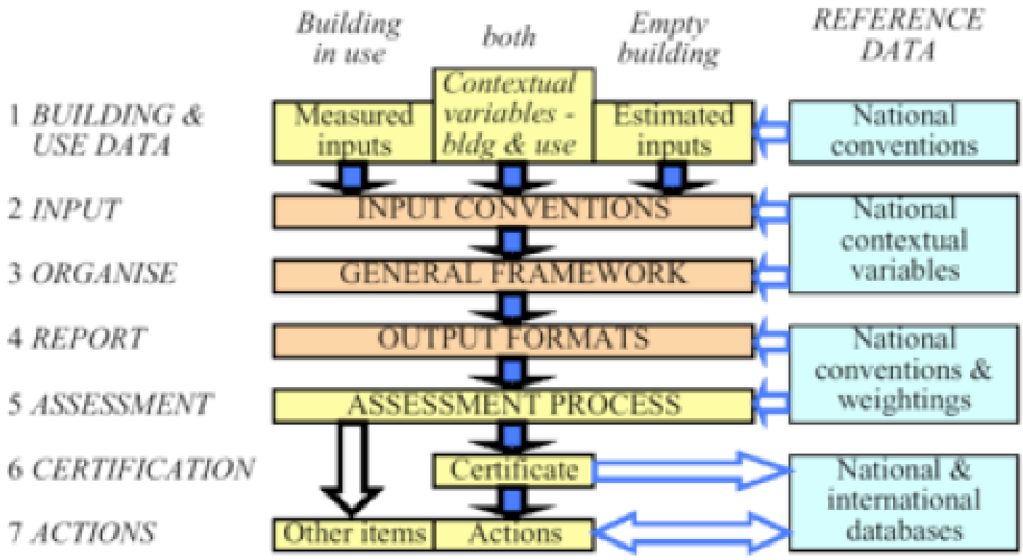
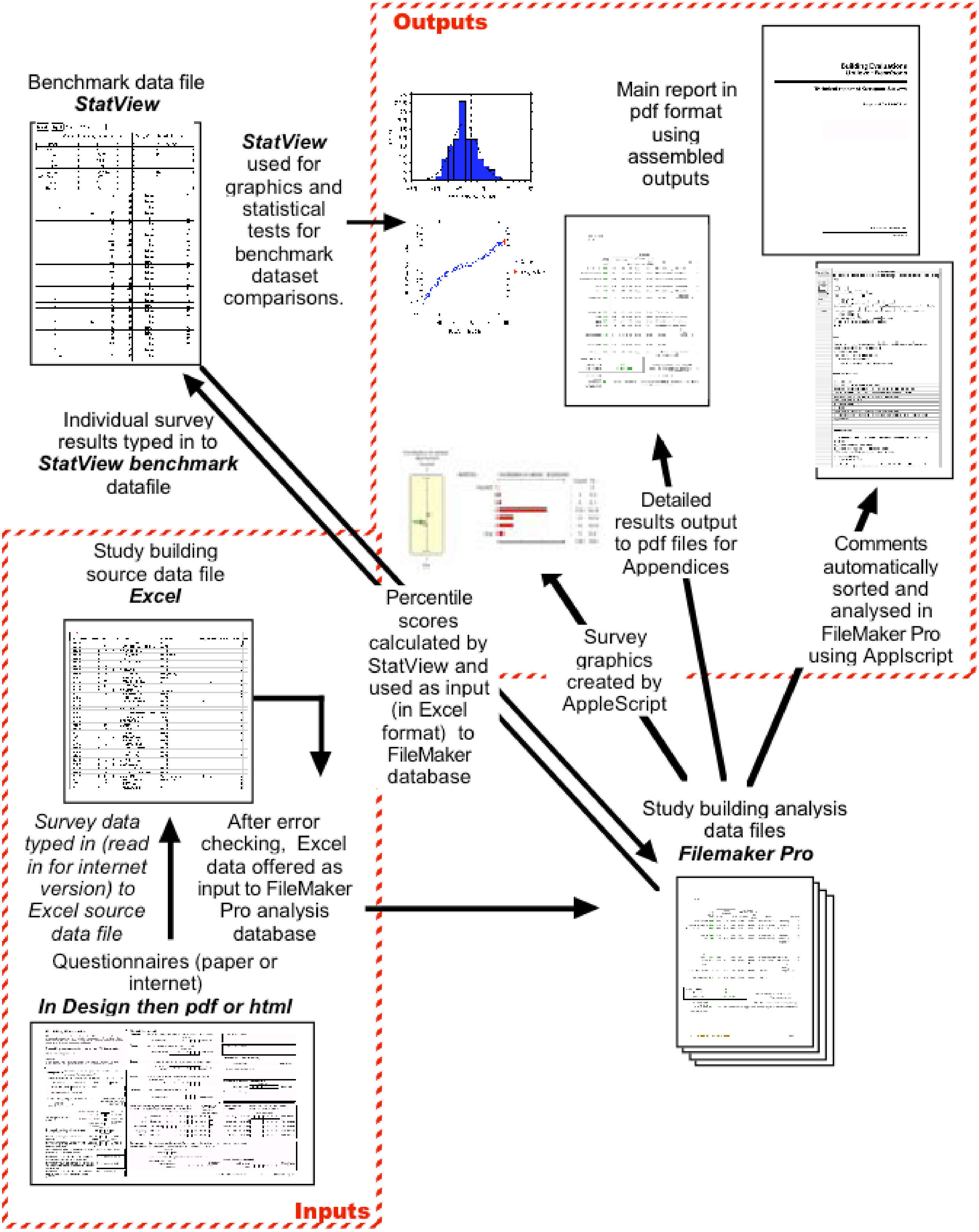


Figure 2 has some of the behind-the-scenes activities involved in a BUS occupant survey.   It has several stages of progressive data refinement, some carried out automatically (e.g. the preparation of website graphics) and some by hand (e.g. updating of the main benchmark database). The system is deliberately maintained as an open system. We find that an open system assists:

- incremental development, there are small changes to the basic system almost every time a study is carried out;
- error checking and quality control, because anomalies can be more easily spotted when interim files can be inspected;
- updates to the system, which are carried out on a study by study basis, with the last study carried out being effectively the latest version;
- our own level of expertise at developing the software - we use relatively simple and 'friendly' scripting languages (e.g. Applescript) because these are more manageable when resources for development work are unfunded and we cannot afford to buy in software expertise.

The system uses six software packages: StatView [16] for statistics, Excel for data entry and data distribution, InDesign for page setting questionnaires, Acrobat for distributing the results, Applescript for data management and workflow production, and Filemaker Pro for the analysis and variable names databases.

Figure 2: Behind-the-scenes activities in a Building Use Studies occupant survey.



Conclusions

IMPROVING TAKE-UP

The point of all this is to make it more attractive for people to use evaluation techniques routinely. However:

- Many are put off by the research culture. Designers, for instance, are often uncomfortable with statistics, data analysis and survey methods. So it is important to demystify techniques as much as possible, explain how they work clearly, especially the assumptions and terminology, and present the results attractively. Equally, the results should not be trivialised or over-simplified, because people may be misled.
- Performance assessment studies are often carried in dissertations and seldom reach mainstream audiences. When they are written up for academic journals, perfectly good studies tend to be rejected by referees because they often do not fit their stricter academic criteria. Results from studies need to appear more regularly in the professional literature.
- Studies have to meet a wider audience, from the corporate manager who only really wants to know how a building affects the perceived bottom line, through designers who will be more interested in detailed diagnostics of building features, to researchers who may be more theoretical in outlook. Methods and their results must meet all of these requirements. They have to be simple enough not to overwhelm readers with detail yet comprehensive to provide them with the answers they are looking for, and easy to understand yet properly underpinned methodologically.
- Results need to be trustworthy and reliable, so people do not think that studies present only the favourable part of the whole story, leaving out the bad news (in which the greatest part of the scope for rapid improvement actually lies). This is why it is better for quality assured methods to be available to third parties, rather than for designers to carry out studies in-house (but it is also important that designers are directly involved with building performance assessments, so that they also learn at first hand, an argument for licensing).
- Ideally, methods should also be structured to some extent by the main targets and headings (hopefully originally) set out in design briefs, so that it is easier to check whether targets have been met and that nothing important has been left out of the analysis. Building evaluation then becomes a routine part of quality control and continuous learning, rooted in the real world.
- Above all, methods must be seen to be value for money. In the past, post-occupancy work has been seriously under-valued by the construction industry and design professions, to such an extent that few have ever wanted to pay for it. Clients, too, do not see why they should have to pay yet more fees for what they regard as, the construction industry's problem.

It is our job to prove the value of post-occupancy studies so that people are prepared to fund them properly, to streamline the methods so that they provide even greater perceived value for lower cost, and to make them a routine for both the occupier and the building provider. From our experience over the past two decades, the second is rather easier than the first or the third!

Notes

[1] See accompanying conference paper: Bordass W, Derbyshire A, Eley J and Leaman A, Beyond Probe: Making Feedback Routine.

[2] Details of the BUS approach, along with other survey methods in the UK, may be found on the *Feedback Portfolio* www.usablebuildings.co.uk/fp.index.html

[3] The BUS questionnaire has a modular format with sections on e.g. thermal comfort, lighting, noise, health and perceived productivity, These can be added or removed as necessary. The 10 questions used in the one-page internet version are always retained for the building rating system.

[4] BUS has found that it is much better to use a single database encompassing all survey variants, than have different databases for each survey type. This means that each study building effectively uses a subset of the latest version of the database, which is refined and updated with every survey. Each survey effectively becomes the latest version of the analysis database.

[5] Software choice is a major concern. We have found that feature-laden spreadsheets with macro programming capability are excellent in situations where their spreadsheet features may be properly exploited, but are much weaker in other areas like graphics, database functions and statistical analysis functions where they are always outperformed by specialist programs. BUS uses spreadsheets only for data input and transfer purposes. All other software is either purpose-made in house, or uses scripts to create workflows between different packages for e.g. statistical graphics, databases and purpose-designed graphics.

[6] Use of in-house software also means that confidentiality can be guaranteed so that e.g. building names do not appear accidentally on web graphics or filenames.

[7] The TM22 energy assessment method is available from CIBSE.

[8] The Probe Pre-visit questionnaire is available under licence from the authors.

[9] Pressure tests by BRE and BSRIA were included in the second and third series of Probe studies.

[10] A water consumption module was available to Probe from WBA

[11] A journey-to-work module was added to the Probe/BUS occupant survey in 2000.

[12] The Feedback Portfolio (www.usablebuildings.co.uk/fp/index.html) has details on some feedback techniques currently available in the UK.

[13]: Federal Facilities Council (2001), Technical Report 145. Learning more from our buildings: a state of the practice summary of POE, Washington, DC, USA.

[14] XML (Extensible Markup Language) is derived from HTML, but ultimately offers greater functionality, more definitional clarity and better facilities for information management and looks likely to be at the heart of future operating systems.

[15] Europrosper. European Programme for Occupant Satisfaction, Productivity and Environmental Rating. See www.europrosper.org

[16] BUS still uses the now obsolete StatView running under Apple System 9. StatView has been superseded by the multi-platform JMP (for Windows, Linux and Apple OSX).

[17] The TM 22 and Europrosper methods use the macro and programming features of Excel.

References

Bordass W, Derbyshire A, Eley J and Leaman A (2004), Beyond Probe: Making Feedback Routine, Closing The Loop, Post-Occupancy Evaluation: The Next Steps, 29 April - 2 May 2004, Windsor, UK.

Federal Facilities Council (2001), Technical Report 145. Learning more from our buildings: a state of the practice summary of POE, Washington, DC, USA.