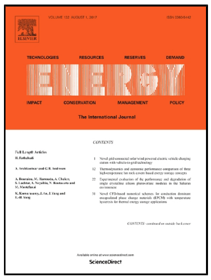
Investigating the consistency and quality of EPC ratings and assessments

Citation for published version:
Jenkins, DP, Simpson, SA & Peacock, A 2017, 'Investigating the consistency and quality of EPC ratings and assessments', *Energy*, vol. 138, pp. 480-489. https://doi.org/10.1016/j.energy.2017.07.105

Digital Object Identifier (DOI):
10.1016/j.energy.2017.07.105

Published In:
Energy



Investigating the consistency and quality of EPC ratings and assessments

David Jenkins, Sophie Simpson, Andrew Peacock

PII: S0360-5442(17)31282-3

DOI: 10.1016/j.energy.2017.07.105

Reference: EGY 11293

To appear in: *Energy*

Received Date: 14 November 2016

Revised Date: 04 May 2017

Accepted Date: 15 July 2017

Please cite this article as: David Jenkins, Sophie Simpson, Andrew Peacock, Investigating the consistency and quality of EPC ratings and assessments, *Energy* (2017), doi: 10.1016/j.energy.2017.07.105

Investigating the consistency and quality of EPC ratings and assessments

David Jenkins[a,*], Sophie Simpson[b] and Andrew Peacock[c]

[a] School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh EH14 4AS, UK D.P.Jenkins@hw.ac.uk

[b] School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh EH14 4AS, UK S.A.Simpson@hw.ac.uk

[c] School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh EH14 4AS, UK A.D.Peacock@hw.ac.uk

*Corresponding author

Abstract

Energy Performance Certificates need a degree of consistency if the information provided by these documents are to be used for applications designed to improve the energy efficiency of the building stock of a country. This study demonstrates, through investigations of small samples of dwellings, that the level of quality, and outputs, from a standardised energy assessment can be variable. Multiple assessors evaluating the same property can produce quite markedly different results and, therefore, recommendations for what that household should do to reduce their energy consumption. Using the results of studies conducted by the authors and from elsewhere, qualified commentary is provided relating to what this form of energy assessment should be used for, and whether improvements can be made in the future as areas of policy look to reflect the recorded energy performance of dwellings through a range of policy vehicles.

Keywords: Energy Compliance; Modelling; EPC; Assessment

1. Introduction

Energy Performance Certificates (EPC), and their accompanying assessment methodologies, are well-established across many European countries as standardised methods for summarising the energy efficiency of individual buildings. Using relatively simple inputs, and basic building physics, an estimation is provided that attempts to grade the efficiency of a property, assuming a “typical” operation within that building.

Emanating from the 2003 European Performance of Buildings Directive (EPBD) (EU, 2010), EPCs allow for a large number of buildings to be assessed across a country in a similar way, thus producing stock-wide results that can indicate a change of energy efficiency over time. They are essentially a means of energy compliance checking, rather than being detailed and representative energy models, using a form of “steady-state” modelling that intentionally simplifies many aspects of thermodynamics. However, more recently, they have also been proposed, and utilised, for a range of other applications that will be discussed in this paper. Before using EPCs for such applications, it is important to understand the background and methodology for any related energy assessment in order to judge the suitability of using an EPC output to make any decision relating to a real building.

In the UK, residential EPCs have recently been used as part of the Green Deal scheme (UK Government, 2015). This allowed householders to apply for loans to cover the capital costs of energy efficiency measures, with an assumption that the energy savings would be able to cover the cost of the loan repayments at a provided interest rate. This assumption was based on savings calculated from a version of the model used with EPC assessments, albeit with several important amendments (discussed later). EPCs have also been proposed for use as an energy efficiency metric with other policy ideas; linking council tax or stamp duty to energy efficiency of the property, gauging whether a property should be lettable (UKGBC, 2013) or linked to the level of feed-in-tariffs that are accessible for a household wishing to install domestic solar photovoltaic panels (Energy Saving Trust, 2015).

A question therefore emerges around what EPCs should be used for. What information do EPCs reliably tell us, and therefore how and where should EPC-related outputs be applied? To answer such questions, it is important to understand both the EPC methodology and the EPC assessment itself (and the individuals who carry out the latter). This paper therefore takes empirical results collected by the authors, as part of a larger project looking at the UK Government's Green Deal, and investigates the quality and consistency of specific types of energy assessments for a small sample of homes.

This particular data was collected as part of a "Mystery Shopper" experiment (DECC, 2014) where multiple assessments were carried out for each home within the sample, with results compared. In doing so, the aim of the paper is to investigate the emergence of inconsistencies within the UK energy assessment approach, and whether such warning signs provide guidance as to how we should formulate future assessment vehicles. It is important to note that this is not primarily a study about "Performance Gap" between models and reality (mentioned in section 2.2); rather, the concern is with consistency and quality, which are more difficult metrics to define. Consistency refers to similarity in the outputs of multiple assessments of the same homes. Quality refers to the visible basis of those recommendations and whether there is evidence that procedure has been followed correctly. Other studies from outside the UK are included as part of a short meta-analysis to provide additional context. This brings together a small number of studies and thus highlights limitations of using simple energy assessment results for a range of decision-making around energy efficiency. Suggestions for a different approach to energy assessment are also provided, based on the above results and the authors' knowledge of building modelling and assessment. The future role of EPCs, and EPC-related assessments, are therefore discussed, highlighting potential knowledge gaps between areas of building physics and energy policy.

2. Background to EPCs

2.1 Standardising energy assessments

The EPBD requires EU member states to apply a standardised calculation methodology for assessing the energy efficiency performance of individual buildings within that country – though differences exist in the detail behind the chosen methodology for each country. Whilst the Directive does propose features of energy performance that should be accounted for (e.g. thermal characteristics of the building envelope, ventilation etc (BRE, 2006)), these are left relatively vague and therefore provides scope for considerable differences between countries. Linked to this is the requirement for minimum energy performance thresholds for new buildings, and EPCs for buildings when they are built, sold or let.

The UK uses the Standard Assessment Procedure (SAP) (BRE, 2012) to meet this objective. A Reduced Data version of SAP (RdSAP) (BRE, 2015) is used for existing buildings, to guide the assessors towards suitable input data for running the building model (e.g. proposing wall U-values for homes of a given age). This accounts for the fact that accurate building input data is more difficult to specify for existing buildings, where plans and drawings might not exist.

This level of standardisation should allow for a degree of consistency in terms of the EPC ratings provided for near-identical dwellings; this is particularly the case as occupancy is based on an assumed “typical” description for that dwelling and linked to floor area.

2.2 EPCs and the Green Deal in the UK

Pay-As-You-Save (PAYS) schemes can differ depending on application, but the central tenet is that the initial capital cost of a (usually) refurbishment technology is repaid by the savings accrued by that technology over a period of time. A consumer can therefore borrow the capital from a lender in the belief that this technology will reduce their energy bills to cover such loan repayments. The onus placed on modelling here is quite clear; rather than the occupant being happy with any type of saving from a given energy-efficiency measure, they want the savings to be enough to repay the cost of that measure, over a period that should be significantly less than the expected lifetime of the measure. This should also include, if they exist, interest payments on the loan. Specific payback periods, and estimated annual savings, might therefore be viewed as a requirement for making that technology economically viable, rather than being used as a rough guide for how much it might improve the energy efficiency of a building.

The well-documented Green Deal in Great Britain, running from 2013 to 2015, was such a scheme. The capital cost of a measure was paid through a loan from a Green Deal Provider, and this was then repaid through monthly repayments by the householder (though it was also briefly available to businesses also). The size of repayments would be less than the refurbishment energy bill savings estimated from the RdSAP model, as discussed in section 2.1, though the RdSAP modelled outputs were amended using an Occupancy Assessment (see below). With an interest rate included on that loaned finance, the hope was that there would be an incentive for both the recipient of the finance (reducing their energy bills) and provider of that finance (through interest rate payments over periods of time up to 25 years) to be involved in the scheme. The scheme was also linked with a supplier obligation scheme, the Energy Company Obligation (ECO), providing payments for certain technologies in certain circumstances (Ofgem, 2015), thus reducing the amount of money that had to be secured through a Green Deal loan.

Green Deal assessments ended up as, essentially, appendices to existing SAP-based energy assessment methodologies. An EPC-type assessment would be carried out (with a standard EPC lodged in the Green Deal Assessment Registration Database (GDORB, 2016) for that dwelling) but this would then be amended by a series of additional questions to tailor the occupancy patterns assumed in an EPC, and some technology/appliance factors, to something more specific to the assessed dwelling. This amended process is often referred to as the Green Deal Occupancy Assessment, abbreviated in this paper as the OA. The Green Deal Assessment report could therefore have quite different recommendations than the matching EPC, as they would be based on slightly different assumptions about the household itself (though the physical description of the house would be near-identical). This additional information from the OA could range from numbers of fridge/freezers, cookers, showers taken per day and real energy bill statements (if the household had these available during the site visit). It is likely that such additional information would make the assessment less standardised, as there would be more potential for disagreement between two assessments of the same house. A valid hypothesis, briefly mentioned in section 3, might therefore be that a Green Deal Assessment would be less consistent than an EPC in terms of replicability of results for a given house.

The uptake for this scheme was generally disappointing, with only 8,300 Green Deal plans in the first two years of the scheme (from beginning of 2013 to end of 2014), despite there being 447,000 Green Deal assessments over this time (DECC, 2015). There are many reasons for this low conversion rate, which goes beyond the assessment procedure and building modelling (such as suitability and attractiveness of the Green Deal loan as a financial product). However, it could be argued that the role of building modelling (and the effect of known model limitations) is often understated in critiques of this scheme, and similar PAYS approaches.

It is very difficult to place a quantified error margin on the accuracy of the RdSAP model in estimating energy bills, and it would probably be unfair to do so as modelled outputs of this nature should not be seen as predictions of, or replacements for, empirical data. However, the "performance gap" between real and modelled energy performances of homes (and other buildings) are well known (Affinity Sutton, 2013; de Wilde, 2014; Visscher 2016) and so it is quite possible for the predictions of energy savings (which help gauge the size of loan repayments) to be quite different to the savings actually experienced by the occupants of a refurbished home. In simple terms, the combined energy cost for a refurbished dwelling (of energy bill plus Green Deal repayment) could be higher than that of the pre-refurbished dwelling (just energy bill).

Whether such an event is likely to happen cannot be discerned accurately by a Green Deal assessor of the dwelling, though safeguards are in place to reduce the chances of this happening (for example, by making the assumed savings from certain technologies more conservative through "in-use factors” (DECC, 2012)). However, such amendments (and, indeed, all changes to the RdSAP methodology that were used for Green Deal Assessments) were never really proven to produce more accurate or reliable results than the simpler EPC assessment.

The registering of the EPC assessments, carried out as part of the Green Deal, does allow for a comparison of consistency between the EPC part and Green Deal part of an assessment. They also provide an opportunity (as discussed below), to investigate the EPCs that were generated by Green Deal assessors who may have come through a slightly different training regime than assessors who were active prior to the commencement of the Green Deal.

3. Quality of assessments and EPC outputs

There is still limited, independent empirical data on the quality and consistency of EPC-based energy assessments for homes in the UK and elsewhere, though several studies of relatively small samples begin to describe areas of concern that, at the very least, require further research. The focus for this paper is on existing work carried out by the authors as part of a project for the UK Department of Energy and Climate Change (section 3.1), but the results of other studies are also described to assess whether common messages emerge that could be acted on by those developing policy linked to uses of EPCs.

3.1 Green Deal Mystery Shopper experiment

3.1.1 Methodology behind field trial

The main study of this paper was part of a larger project (DECC, 2014) looking at the customer journey and assessment process behind the Green Deal, project managed by ICF-GHK. As part of an analysis of the energy assessments themselves, a “Mystery Shopper” study was conducted for 29 volunteer households (NB more signed up for the exercise but were unable to find enough energy assessors to qualify for the full analysis). This entailed the householder finding (at least) four independent assessors to carry out a full Green Deal assessment of their house, where each assessor would assume that they were carrying out a standard assessment for the customer (though, as required for such exercises, a general announcement was published that such a study would be taking place over a given period of time). A fifth “reference” assessment was provided by the organisation C A Design Services (CADS) Ltd, who were part of the project team and were therefore able to return to a property after the four assessments to carry out any additional investigation as to why there may have been anomalies. The project team also analysed the entire customer journey, from the first phone call to the report and recommendations provided after the site visit; however, this is not part of the paper (but the methodology can be found in the aforementioned DECC report). This paper will therefore focus on the energy assessment results, particularly those relating to the EPC part of the assessment (though results of the Green Deal OA amendments will be discussed where relevant).

The full dataset therefore included 145 Green Deal assessments (and accompanying EPCs) across the 29 sample dwellings. Due to the geographic spread of the homes across England and Wales, no assessor would have assessed more than one home. Although this represented a small sample (if compared to, say, the various UK house condition surveys), it did provide enough assessments to identify inconsistencies in approaches to energy assessments, and highlighted disagreements in outputs for a given home.The small sample does mean that the quantified results provided below cannot be statistically extrapolated to the entire housing stock; but it is likely that issues found within this sample are not just confined to the sample.

3.1.2 Variability in EPC ratings

While these assessments would have been driven by the Green Deal and checking for eligibility for Green Deal loans, standard EPCs were also lodged to the assessment database and so can therefore be analysed (using reference identifiers that link Green Deal reports with respective EPCs emanating from the same site visit). The EPC results collected from this database are shown in Figure 1 for the 29 sample dwellings. The horizontal lines on each data point show minimum and maximum ratings for that property (across the four assessments), a mid-data point showing median value of the four assessments, and the CADS value also shown as a reference. Significant variation was observed across this small sample, even though assessors would be tasked with using the same methods, models and procedures. It can be seen that almost two-thirds of the 29 dwellings had assessments varying by at least two energy performance bands, with two dwellings having assessments spanning three bands. The average range (difference between maximum and minimum) in EPC rating value across all 29 dwellings is 11.1.

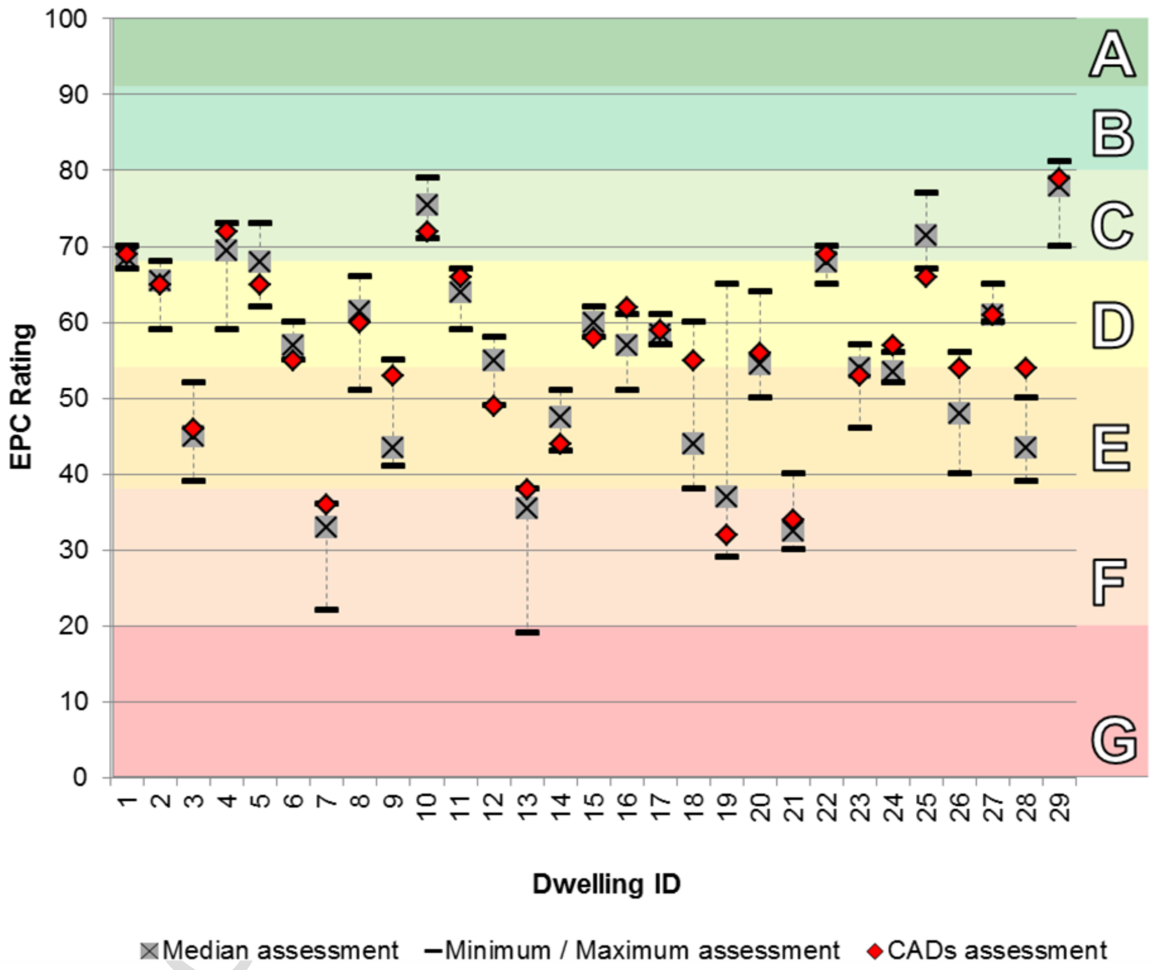


Figure 1 – EPC ratings calculated by multiple assessments of 29 anonymised sample dwellings

Along with the EPC results themselves, the database records some of the inputs used in the assessment by the assessors; however, the input sheets used in each RdSAP model were not accessible as this would have required communication with the individual assessors (negating the "Mystery Shopper" aspect of the work). Therefore, whilst some inputs are clearly recorded (e.g. assumed floor area), others are more opaque. For example, building fabric options (and some lighting and heating technologies) were defined by an output of the EPC register database that categorises options by a five point energy efficiency rating, where 1 is poor and 5 is very good. Quantifying the impact of some building fabric inputs was therefore often not possible from the available data. Specifically, the thermal performance of a wall in the RdSAP model is defined by the type of construction and age of dwelling, which can produce 31 distinct U-values (for England). These values were not visible in the database accessed by the authors. Some more basic inputs, such as property age, were provided by the householders themselves as part of a survey. Thus, input and output data could be reviewed to help describe patterns in the differences reported in Figure 1 with, again, limitations of sample sizes within any sub-category of dwelling.

3.1.3 Investigating variation in EPC inputs

The reported age, construction and property type were investigated as possible indicators for why assessors may have disagreed on the thermal properties of these dwellings. Isolating such variables within an assessment vehicle that has many dozens of variables, some of which are not recorded directly in the database (as well as mistakes and misjudgements, possibly, being part of any discrepancies), is not possible. Using the results with some knowledge of how the assessment (and RdSAP model) works can still be useful, as will be discussed.

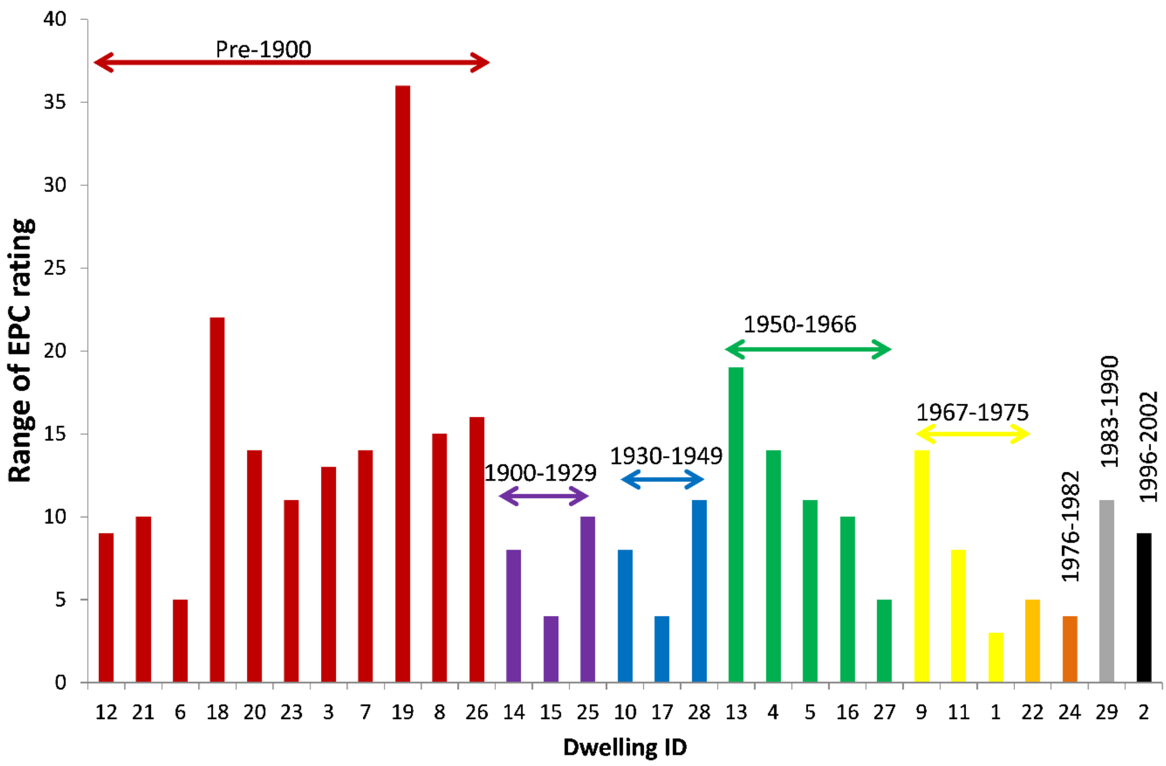


Figure 2 – Range of EPC rating for every dwelling ordered by occupant-reported age of dwelling

Figure 2 takes the ranges of Figure 1 and re-orders them by property age category, using the age band of the property as reported by the occupant. Due to the small, and variable, numbers of dwellings within each sub-sample of age group, a direct comparison comes with clear caveats, though some dwelling types appear to show more inconsistency than others. For example, some pre-1900 dwellings appear to have particularly high ranges in estimated EPC rating. If a larger sample was available, property type could also be factored into this analysis to indicate causes of inconsistencies. There is a suggestion of the importance of house type from dwelling no.19 (which returns the highest range), which is described as a flat/maisonette/tenement. These dwellings could be more likely to exhibit differences in modelling approaches as a result of assessors needing to make decisions about, for example, whether an internal wall is adjoined to a heated space from another property or not. Dwelling no.4 is also the same property type but does not seem to suffer from the same level of inconsistency in the energy assessments; however, this is a more modern brick-built property as opposed to the stone construction of no.19 (again suggesting that knowledge of both property age and property type might be important for future studies of larger samples).

One possibility is that older properties are not well specified in SAP-based models due in part to difficulties in defining the heat transfer properties of the building fabric. It is possible that the users of the models also have difficulties with these dwellings, identifying where properties might have undergone a refurbishment or extension since being built (which becomes more likely as the property gets older).

Total floor area (TFA) is another key parameter that could have an impact on final EPC assessment. The assessor can record this as either “external” or “internal” TFA, with the former including wall thickness but requiring conversion to internal before being used by the RdSAP calculation. Figure 3 shows the variation in the measured floor areas used by the assessors across all 29 dwellings.

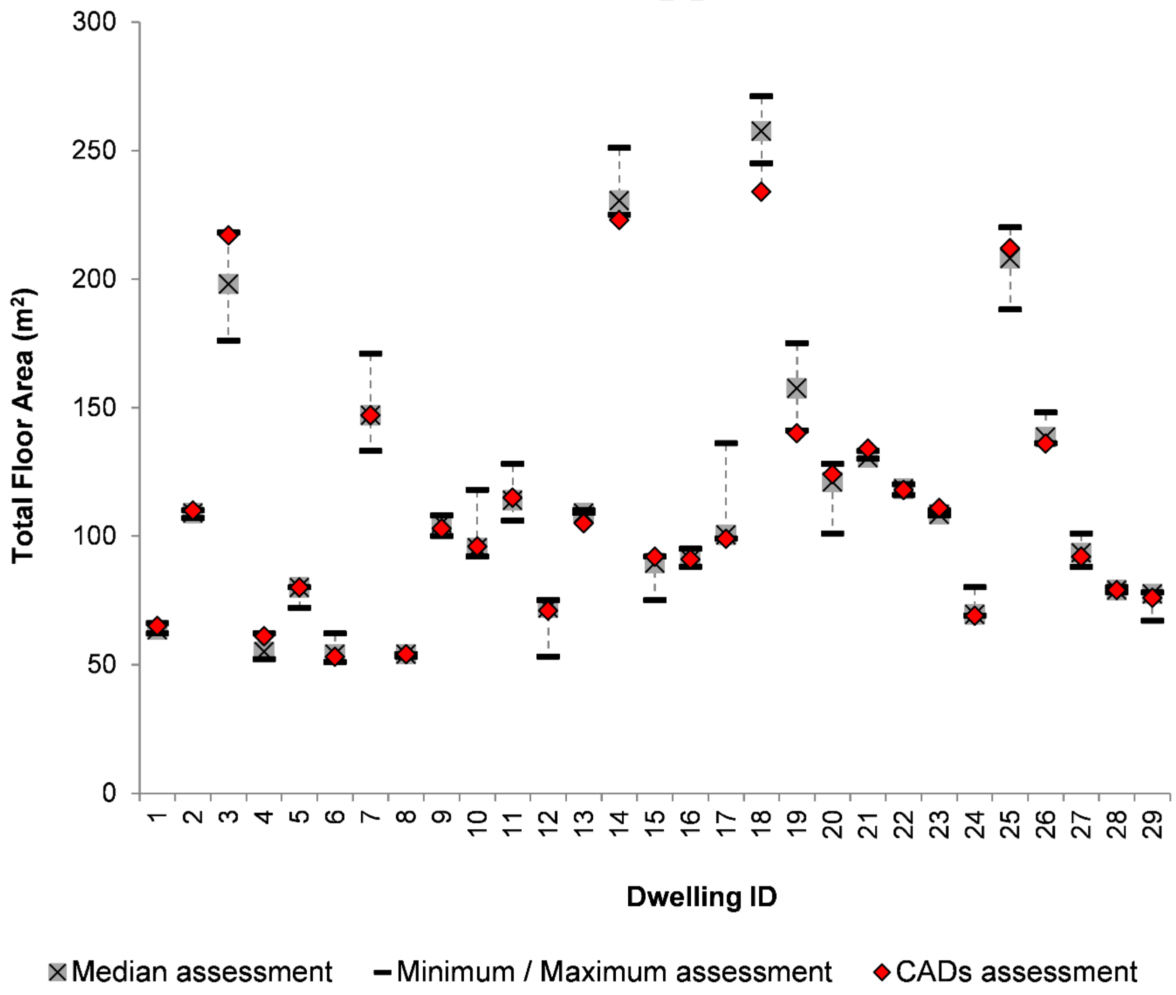


Figure 3 – Values of total floor area for the 29 dwellings with four assessments

Significant variation in TFA was apparent across many properties. Some of this variation is unlikely to be just a consequence of acceptable experimental error during a site visit, such as with measurements of smaller areas within a dwelling, obstacles/furniture inhibiting access, and/or rounding errors.

This data can be reinterpreted as percentage variation in floor area, as shown in Figure 4. This relative metric ensures that variations in larger dwellings are not exaggerated. The average range between minimum and maximum TFA across all 29 dwellings was 13.7% of floor area, though this is weighted towards a smaller value by properties where a very high degree of consistency was achieved (e.g. seven homes had assessments that were within a 5m^2 range, suggesting consistency of floor area input is possible for some homes).

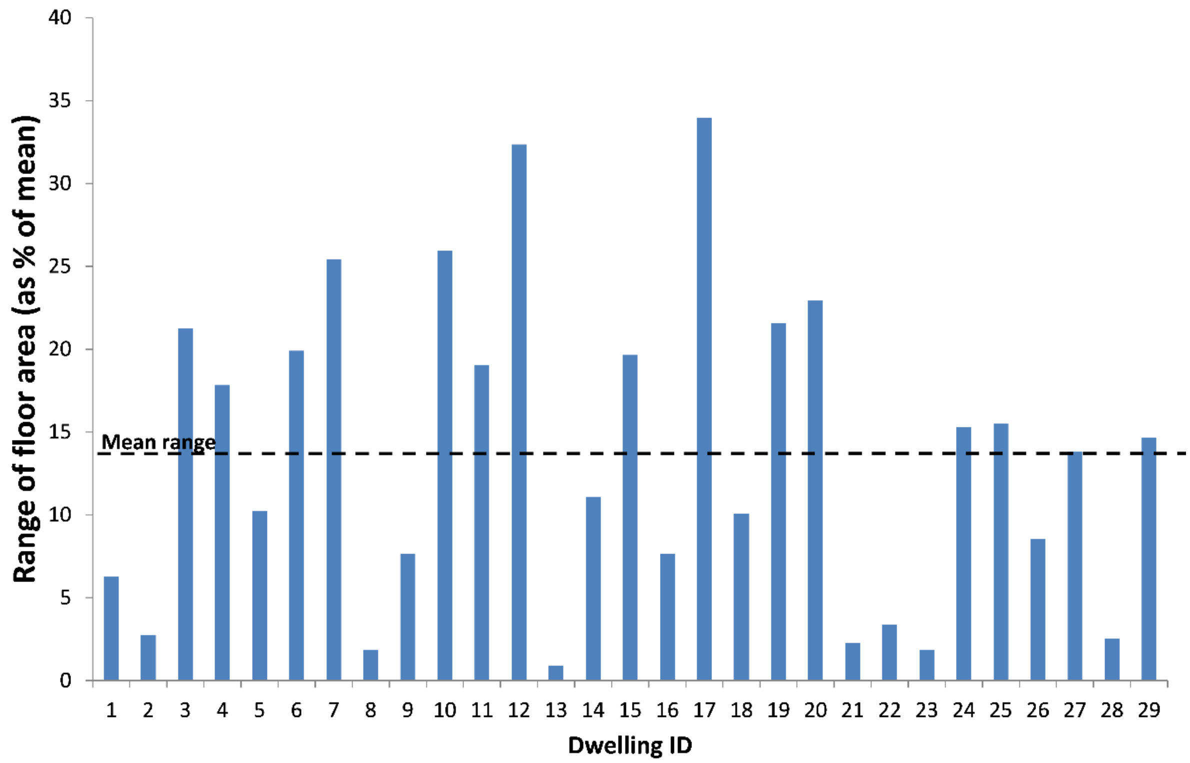


Figure 4 – Range of total floor area as a percentage of the average floor area returned for every dwelling

Perhaps counter-intuitively, dwellings with a high TFA variation were not necessarily those with a high variability in EPC rating and vice versa (e.g. dwelling no.13 has excellent agreement on TFA across all four assessments but a 19 point range in the EPC rating in Figure 1). This suggests that mistakes/disagreements in TFA are probably not the main cause of assessment consistencies (although they may play a role in some dwellings). This is further demonstrated in Figure 5, showing the lack of correlation between variations in TFA (across the four assessments of a given dwelling) and the range of EPC ratings provided for that dwelling.

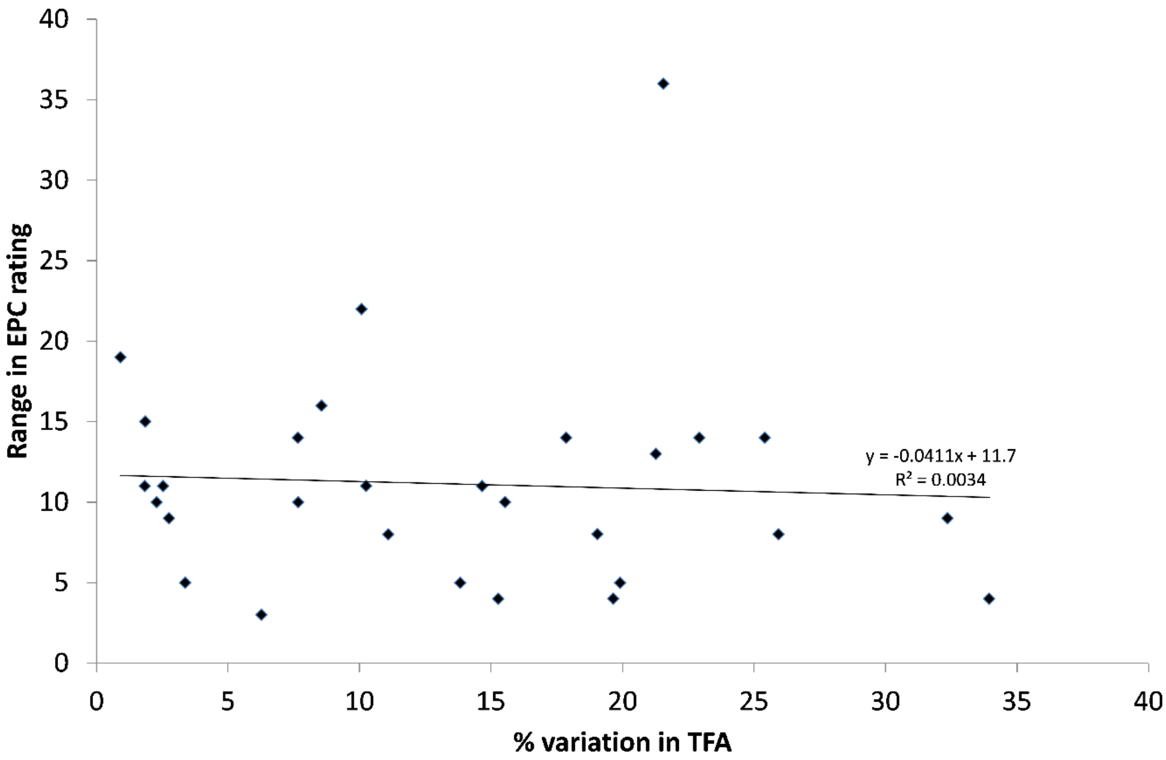


Figure 5 – Correlation between variation in TFA and range of EPC ratings from 29 dwellings

During a feedback process with the CADS reference assessors, it was suggested that aspects of specific houses might have contributed to assessors making different calculations of floor area. These aspects included the existence of a "room-in-roof", whether heated conservatories or porches were included in floor area, and likewise non-heated basements. However, it was suggested during this feedback exercise that the most likely cause was basic individual error (though such a judgement is, of course, subjective).

Even accounting for the previously described problem of the use of proxy variables for some inputs (like "energy efficiency rating" of building fabric, rather than specific U-values), it was still possible to carry a certain level of input analysis across the different assessments. Figure 6 demonstrates, for some chosen parameters, the number of dwellings where all four assessors were in agreement. The least consistent parameter out of this selection appears to be roof energy efficiency rating, with only six dwellings given the same rating by all four Green Deal assessors of that property. The fifth assessment undertaken by the CADS assessor suggested that some variations could have been attributable to an assessor not accessing the loft space and therefore recording a default poor rating [NB It was noted, in the householder feedback survey, that some lofts were not accessed by all assessors for that dwelling – suggesting that accessible lofts were sometimes ignored]. There may also have been disagreements around the effect of boarding and/or compression of insulation in loft spaces.

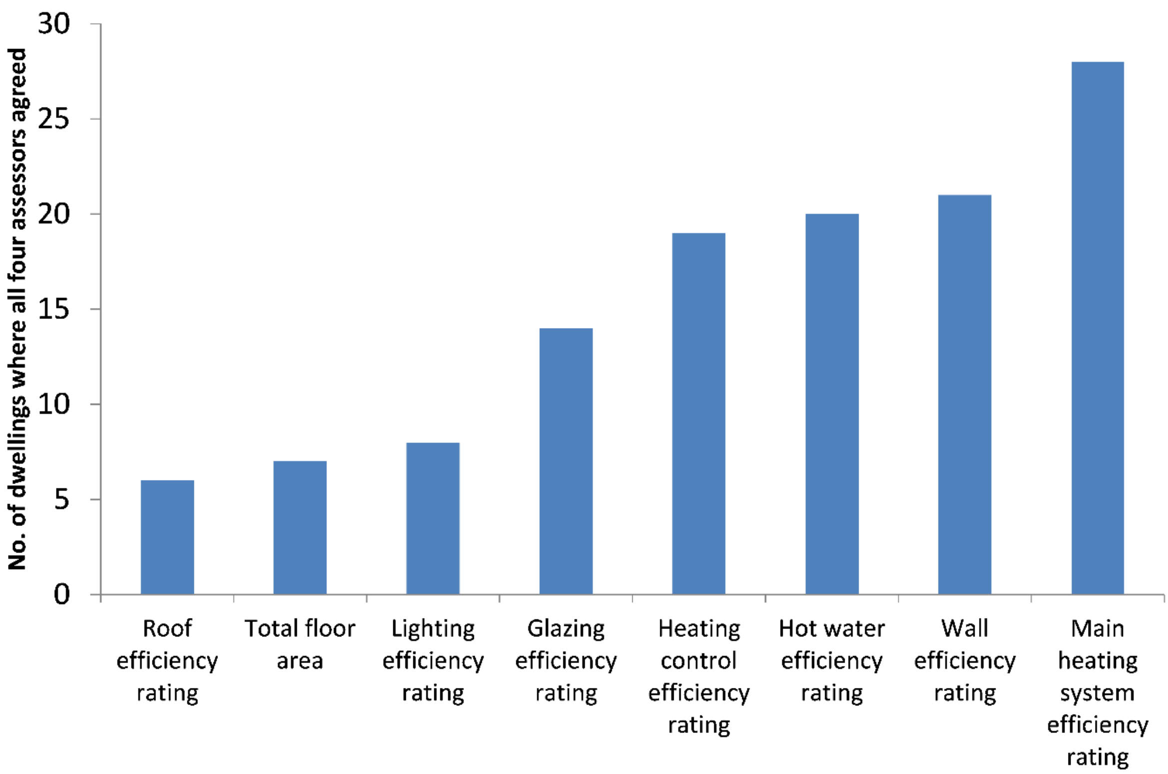


Figure 6 – Number of dwellings where all four assessors agreed on parameters used in the EPC calculation process

3.1.4 Calculated EPC Space and Water Heating requirements

Estimations of space and water heating consumption are calculated from both the EPC and OA parts of the assessment. The values within the EPC calculation principally rely on standardised assumptions and therefore, should, provide less scope for variation among assessments than might be observed with any equivalent OA values. Figure 7 shows the significant variation in calculated space heating costs from the EPCs across the four assessments, with the fifth reference assessment included separately. A difference of more than £150 in estimated annual space heating costs is seen for 23 of the 29 dwellings, with greater variation seen for properties with higher space heating values. Also, if assuming that the fifth reference assessments are more likely to be "correct", it is not necessarily the case that variability is coming from just one anomalous assessment: for dwellings no.7, no.9 and no.18 the retained CADS assessor's results are closer to one of the outlier values.

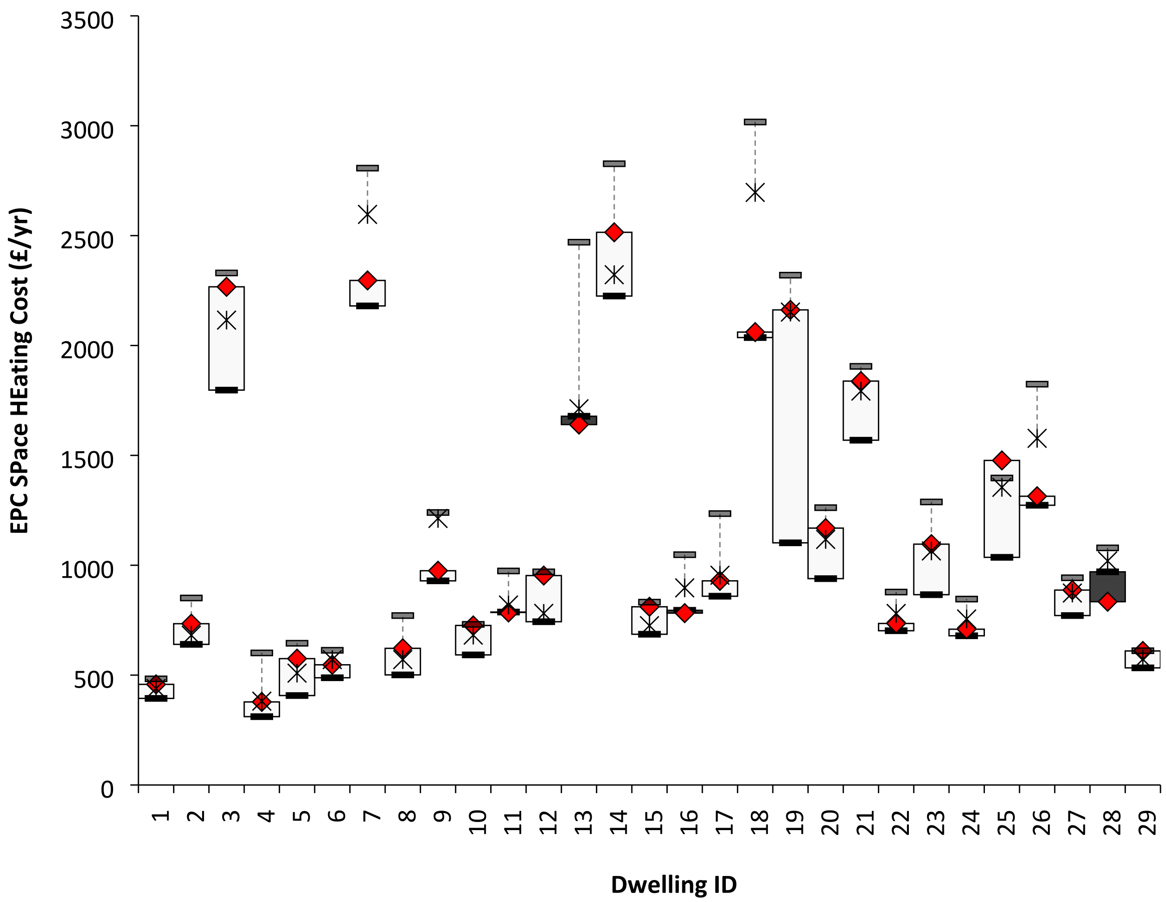


Figure 7 – Calculated annual space heating costs from EPCs of 29 dwellings with four assessments

For the 29 assessed homes, space heating accounts for the greatest proportion of the annual energy costs calculated by the EPC – as might be expected from empirical data also for a UK home. A discovered correlation between variability in EPC rating (from Figure 1) and variability in EPC space heating cost (with a linear correlation coefficient of 0.83) was therefore not surprising. This is despite the fact that EPC rating is independent of floor area (i.e. the annual energy use is evaluated on a per square metre basis). One result of this is that, for a given dwelling, an assessor reporting a higher floor area is more likely to determine a lower annual space heating cost per unit area, which will produce a lower EPC rating.

Across all 29 dwellings, the mean range in space heating cost is £355, which is significant if compared to the UK's average annual gas bill per household of £493 (from the 2013 UK housing energy data (DECC, 2013)). There are also examples of dwellings well in excess of this mean range: for dwellings no.13, no.18 and no. 19 there were ranges of £830, £956 and £1,218 recorded respectively.

Variation can also be observed in Figure 8 for the estimated annual hot water costs. For RdSAP modelling, this is largely determined by the number of occupants which, in turn, is a function of floor area for the EPC calculation and the specification of the hot water system (e.g. heat generator type, efficiency, and storage). However, the influence of floor area variations on hot water variations seems to be relatively insignificant. For example, there is a considerable range in hot water costs for dwellings no.21 and no.28, despite both of these having assessors that generally agreed with each other on floor area (Figure 4). This indicates that that a large degree of the hot water consumption variability is provided by differences in hot water system specification, or individual errors from the assessors.

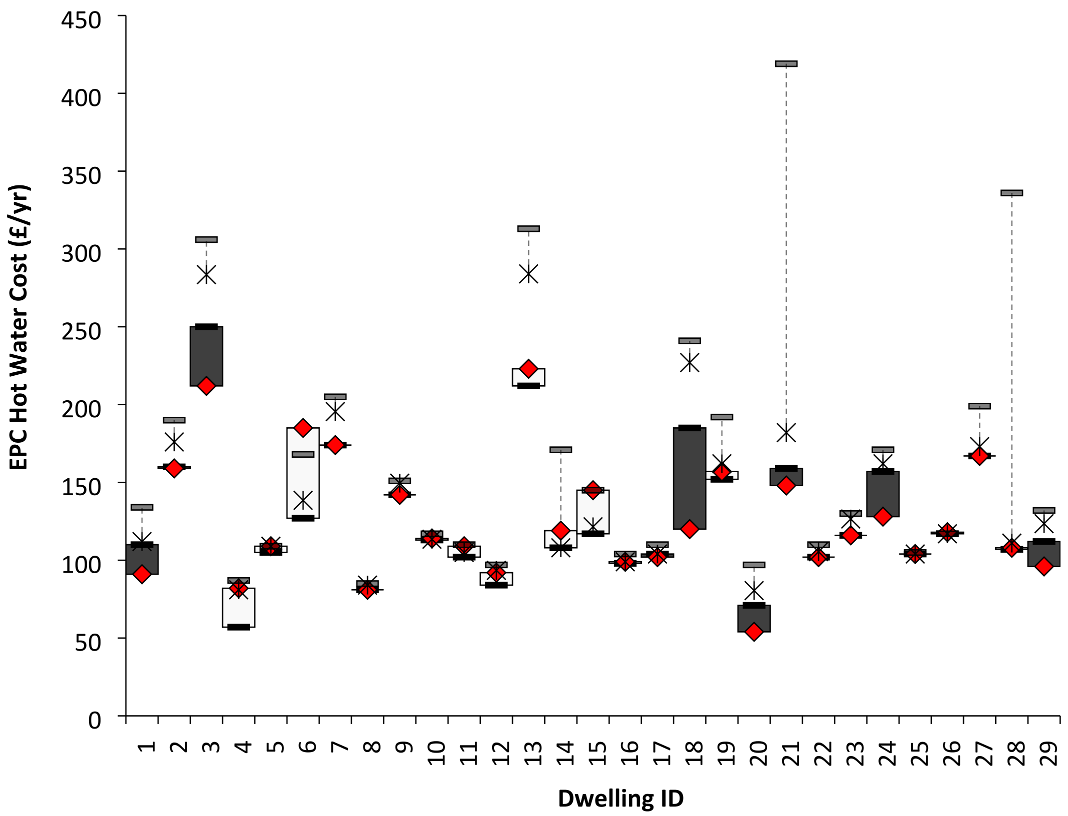


Figure 8 – Calculated annual domestic hot water costs from EPCs of 29 dwellings with four assessments

3.1.5 Further inconsistencies for the Green Deal

Although not the focus of this paper, the aforementioned report (DECC, 2014) identifies even greater inconsistencies and disagreements when the analysis is continued for the Green Deal assessments. For example, Figure 9 shows the lack of agreement between assessors when identifying key inputs for the OA. These would not have directly affected the EPC stage of the assessment, and are arguably subject to individual judgement in a way some EPC inputs are not, but they further highlight the different approaches that assessors were conducting when collating this information. The Green Deal Assessment Reports that these households received therefore showed very large disagreements in the types and numbers of recommendations proposed, as explored in the full report. This is likely to be due to the aggregation of multiple input disagreements from both the EPC and OA stages of the assessment.

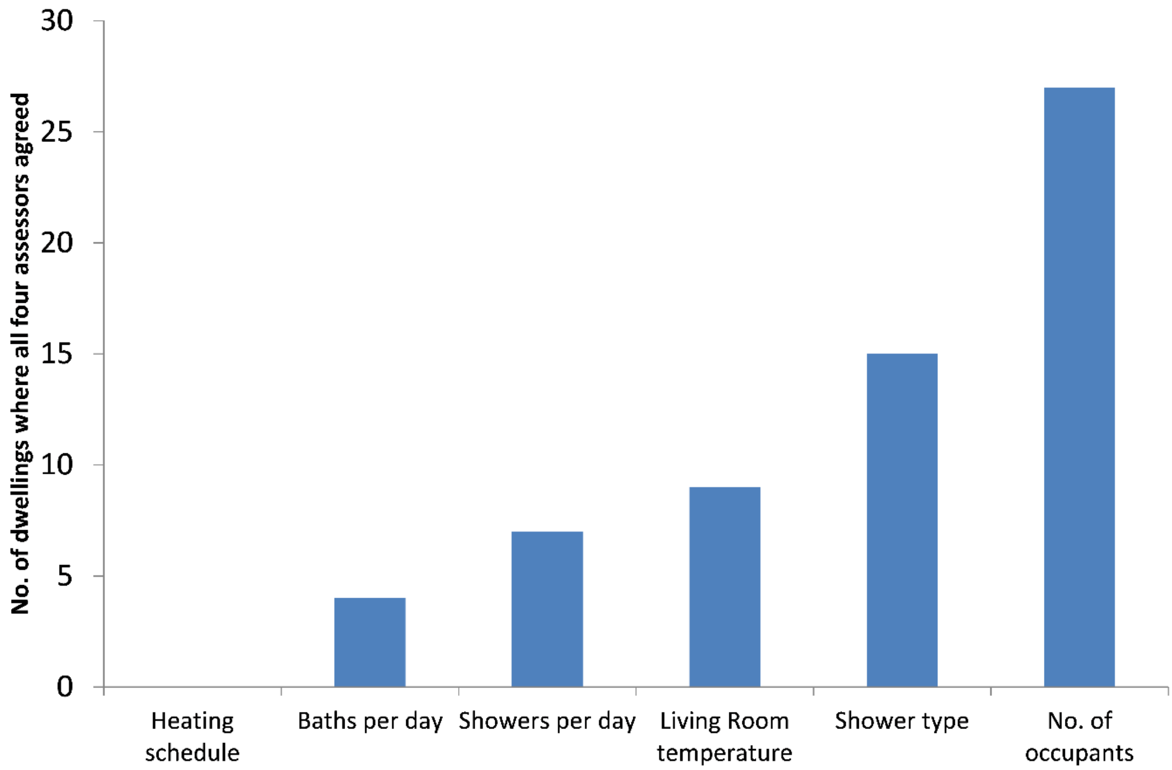


Figure 9: Number of dwellings where all four assessors agreed on parameters used in the Green Deal OA calculation

3.2 Other studies of energy assessment quality

For studies aiming to understand the detailed input (and output) from assessments of individual dwellings, datasets corresponding to relatively small samples can be the result. This is again the case for a study in Italy (Tronchin and Fabbri, 2012) which took a single detached house that was assessed by 162 independent trainee assessors. Clearly, it is impossible to generalise the results to other buildings, but as a case-study it is of interest. For this study, the authors assumed a “correct” value of the building, trusting that an initial energy assessment that produced an EPC rating of D was generated in the correct way. In reality, two assessments that (in isolation) are seemingly correct could still produce different answers if the methodology allows for a degree of choice by the assessor (as discussed by the aforementioned Mystery Shopper work).

Mistakes were observed with some of the 162 assessments, along with differences in assumptions behind, amongst other factors, thermal bridging, boiler efficiency, geometrical data and domestic hot water production. The standard deviation across all assessments for the presented energy index (total annual energy consumption per unit floor area) was 22.5 kWh/m^2/year; for an average consumption of 106 kWh/m^2/year, this is quite large. Moving away from the numerically modelled energy consumption, 72% of assessments agreed the house was D rated, 14% thought C rated and 9% E rated. Six assessments suggested the house was B rated and one as high as an A rating, though these were explained as being mistakes by the assessors, rather than more subtle differences in assumptions applied to the modelling. The authors seem to suggest, with some justification for EPC assessments, that the most important factor is an agreement on the energy rating of that dwelling, and therefore 72% assessments are deemed to meet some measure of success. This still leaves over a quarter of assessments of a relatively simple building being, by this measure, “wrong”.

In the UK, a report by the Zero Carbon Hub (Zero Carbon Hub, 2014) identified a series of areas that might result in poor quality energy assessments for new-build projects. Issues of concern included: a lack of accurate input information from the design specification; accuracy and lack of validation of the SAP model and inability to account for dynamic effects; communication between different individuals within the design process; the use of as-designed rather than in-built data (a problem also identified elsewhere (Pan and Garmston, 2012)) and several issues about the use of representative input data. For this last point, the report carries out a survey of SAP assessors to query where information is lacking when calculating an EPC for a new-build property. It was found that air-tightness testing and space heating details are usually provided for such assessments (in the region of 95% and 88% respectively are recorded for whether this information is either always provided or only occasionally missing). However, SAP checklists for ventilation (16%), as-built drawings (30%) and U-value calculations (40%) are far less commonly used.

Competency of SAP assessors are also questioned in the above report, which found evidence of errors relating to thermal bridging and U-value calculations. Quoting research elsewhere, for a study looking at SAP assessments of 82 dwellings (DCLG, 2009), errors were identified in the recording of wall areas, storey height, other building element areas, orientation and floor area. This is despite the fact that, being new build, such information should have been readily available. It is therefore perhaps unsurprising that similar areas of error were recorded for the existing dwellings of the Mystery Shopper exercise described in section 3.1; though, once again, care must be taken when extrapolating such conclusions to outside the small sample.

These results are similar to more anecdotal evidence provided by industry (Building.co.uk, 2013), where due diligence of previously performed EPC assessments show a combination of mistakes, misjudgements and questions of whether the EPC is genuinely reflective of the energy being used in that building.

4. Messages emerging from studies

From the main Mystery Shopper study summarised in this paper, and relevant studies referenced, it is difficult to quantify to what extent observed inconsistency emanates from poor guidance from official methodologies or poor understanding (or simple mistakes) from the assessor using that model. For the Mystery Shopper study, it is clear that assessors within this sample had different approaches to conducting an EPC and, if such inconsistency was replicated across all EPCs, the reliability and usefulness of EPCs could be rightly questioned. It is fundamental to a standardised compliance method that all modellers, and assessors, use the models in the same way.

If building owners and businesses are being encouraged to make decisions on property portfolios based on such EPC ratings, it is surely valid to ask whether this approach is appropriate. Likewise, if householders are encouraged to use their own capital to improve their homes based on EPC-related data, validation of the models behind this advice is of great importance. This area requires more research to uncover the causes of such inconsistencies and their impact, as well as a better form of auditing to instil confidence that the assessment framework is fit-for-purpose.

It could be surmised that some of these identified problems of inconsistency are actually a consequence of the "success" of EPCs. Having decided, across many European countries, that EPCs are required to visibly identify levels of energy efficiency in the building stock, our assessment methodologies then need to be carried out for a vast number of buildings, with a similarly vast number of assessors for achieving this. With the observed variations in approaches already documented, it might be reasonable to question the level of training that these assessors enter. The key goal of such training is often to be able to use a model effectively, not necessarily to understand the theory underpinning the model (and the physics of buildings more generally). For example, Green Deal assessor training constituted, typically, a week's training course – though more detailed courses run for two weeks. The impact that this fast-tracking of energy assessors has on assessment quality requires further research.

The modelling community should therefore be aware that compliance models potentially have to be used by people with limited understanding of the calculation engines of those models. In itself, this does not have to be a problem providing the framework for using that model does not allow scope for variation in model outputs for a given property. Quality control and consistency becomes a huge challenge when so many assessors are available for hire within the energy assessment market, and there is some evidence that the current level of validation of these models, and the modelling process, is not sufficient.

5. Proposing alternatives

It is not practical, or likely, that the whole approach to energy assessment will be altered based on the currently identified flaws. Nevertheless, such problems might suggest that now is a good opportunity for reviewing what current EPCs are being used for; and, even with a more consistent and validated methodology, what they should be used for in the future.

A key role of EPCs is in improving energy efficiency of existing buildings – both identifying an approximate baseline of energy use and suggesting tailored measures for reducing that. If we consider individual measures, it is reasonable to question the suitability of an EPC for doing this – both for areas where it appears unnecessary (for low-cost measures that are known to have a high likelihood of being successful), and situations where the model does not have the required calculation engine to capture certain aspects of building physics. Focussing just on dwellings, for example, we might question whether is it necessary to have a full-house energy audit to advise a householder if they need loft, or even cavity-wall, insulation. With the merits of such technologies quite well known, establishing an audit that is more focussed on practical issues (e.g. space in loft, the type of cavity identified and whether it should be filled) might be ultimately a more cost-effective process for encouraging lower cost efficiency improvements. This argument appears even more valid when it has been suggested elsewhere (Stone et al, 2014) that 75% of the variance of energy ratings produced by the SAP method can be accounted for by just considering geometry, heating efficiency and wall U-value. For many applications, an even simpler modelling approach (or no modelling approach at all) might not have an impact on the success of the end goal of increasing the uptake of energy efficiency measures.

For some higher-cost improvements (solid wall insulation, heating technologies, onsite generation), it might be appropriate to question the ability of a steady-state model (like SAP) to model these – and whether the householder receives reliable advice on the effectiveness of such technologies. There may exist new-build homes that would benefit from a standardised dynamic modelling procedure (Marshall et al, 2016) that can, for example, attempt to quantify the impact of thermal mass changes following a solid-wall insulation improvement, occupancy patterns or dynamic effects of heating systems over time.

With SAP-based models using "average" assumptions (particularly for occupancy and technology performance), are these only suitable when applied to forms of stock modelling? It would seem appropriate, from the research of section 3, to at least discuss the value of applying these to individual properties and, if doing so, how to communicate the results to the occupants of those properties in a way that is not disingenuous.

Finally, we are entering an age where improvements in data storage and technologies such as smart meters may enable us to have a vastly improved dataset of empirical energy use in buildings. Formulating semi-empirical models that are calibrated on such data in a statistically valid and standardised way (rather than the approximation used with Green Deal OA where annual energy bills are used in an inconsistent way in an attempt to match modelled energy use with some representation of real data) could be a way forward. This could involve real energy data, or real physical data (Dineen et al, 2015), moving our estimates of building energy efficiency into a more empirical domain.

6. Conclusion and Policy Implications

Accuracy and consistency are two different, but related, issues when investigating the outputs of energy assessments. Much research over the Performance Gap suggests that there are limitations to simple steady-state models achieving accuracy when compared with real energy data. However, if such assessments achieve consistency, through clear methodologies applied in standardised way for all buildings, then it is at least possible to gauge what models are telling us about the building stock, and then assess whether the same messages (even if slightly different values) are being produced from both modelled and empirical data.

If we begin to question whether energy assessments, and resulting EPCs, are as much a product of the individual creating them as any semblance to building physics and occupant behaviour, then there is a clear danger in using such metrics for making policy decisions on energy efficiency.

The study of 29 dwellings proposed here is, in essence, a series of detailed case-studies rather than a representation of a building stock. Despite this caveat, with repeated assessments of each dwelling and additional context from elsewhere, valid concerns can be expressed. Messages emanating from multiple studies, albeit from small samples within each study, have additional resonance when they are consistent with existing concerns over, for example, the ability of a basic steady-state model to reliably predict the energy consumption of even a simple building. The benefits of having a clearly replicable and robust means of estimating energy efficiency levels of our building stock are quite obvious. The strong desire to reach this eventuality should not cloud any attempts to strongly critique the current methods and models, and think about whether other alternatives could still assist our goals of greening the existing building stock, and producing genuinely low-energy new buildings.

Acknowledgements

The Mystery Shopper data analysis used in this work was conducted by the authors as part of a larger project managed by ICF-GHK and funded by the UK Department of Energy and Climate Change. All contributions and assistance from individuals within these organisations are gratefully acknowledged.

References

Affinity Sutton, 2013. Future Fit – Final report

http://www.affinitysutton.com/media/364652/futurefit-quick-links-PDF-1.pdf

Building.co.uk Magazine, 2013. So you thought the accuracy of EPCs was improving?

http://www.building.co.uk/so-you-thought-the-accuracy-of-epcs-was-improving?/5063502.article

Building Research Establishment (BRE), 2006. Energy Performance of Buildings Directive

http://www.bre.co.uk/filelibrary/Scotland/Energy_Performance_of_Buildings_Directive_(EPBD).pdf

Building Research Establishment (BRE), 2012. The Government's Standard Assessment Procedure for Energy Rating of Dwellings https://www.bre.co.uk/filelibrary/SAP/2012/SAP-2012_9-92.pdf

Building Research Establishment (BRE), 2015. RdSAP Conventions for RdSAP 9.92

http://www.bre.co.uk/filelibrary/SAP/2012/RdSAP-Conventions.pdf

UK Department of Communities and Local Government (DCLG), 2009. EEPH/AECOM Research into Compliance with Part L of the Building Regulations for New Homes

UK Department of Energy and Climate Change (DECC), 2012. How the Green Deal will reflect the in-situ performance of energy efficiency measures

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48407/5505-how-the-green-deal-will-reflect-the-insitu-perfor.pdf

UK Department of Energy and Climate Change (DECC), 2013. UK Housing Energy Fact File

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/345141/uk_housing_fact_file_2013.pdf

UK Department of Energy and Climate Change (DECC), 2014. Green Deal Assessment Mystery Shopping Research https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/388197/Green_Deal_Assessment_Mystery_Shopping_FINAL_PUBLISHED.pdf

UK Department of Energy and Climate Change (DECC), 2015. Green Deal and ECO statistics https://www.gov.uk/government/collections/green-deal-and-energy-company-obligation-eco-statistics

de Wilde, P., 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation, Automation in Construction 41, 40-49

Dineen, D., Rogan, F., and O Gallachoir, B.P., 2015. Improved modelling of thermal energy savings potential in the existing residential stock using a newly available data source. Energy 90, 759-767

Energy Saving Trust, 2015. Feed-in-tariffs http://www.energysavingtrust.org.uk/domestic/feed-tariffs-0

European Union Energy Performance of Buildings Directive 2010/31/EU, 2010 http://eur-lex.europa.eu/legal-content/EN/ALL/;ELX_SESSIONID=FZMjThLLzfxmmMCQGp2Y1s2d3TjwtD8QS3pqdkhXZbwqGwlgY9KN!2064651424?uri=CELEX:32010L0031

Green Deal Oversight and Registration Body (GDORB), 2016. GDORB website and database http://gdorb.decc.gov.uk/

Marshall, E., Steinberger, J.K., Dupont, V., and Foxon, T.J., 2016. Combining energy efficiency measure approaches and occupancy patterns in building modelling in the UK residential context, Energy and Buildings, 111, 98-108

Ofgem, 2015. Energy Company Obligation (ECO) https://www.ofgem.gov.uk/environmental-programmes/energy-company-obligation-eco

Pan, W., and Garmston, H., 2012. Compliance with building energy regulations for new-build dwellings, Energy 48, 11-22

Stone, A., Shipworth, D., Biddulph, P., and Oreszczyn, T., 2014. Key factors determining the energy rating of existing English houses, Building Research and Information, 42, 725-738

Tronchin, L., and Fabbri , K., 2012. Energy Performance Certificate of building and confidence interval in assessment: An Italian case study, Energy Policy 48, 176-184

UK Government, 2015. Green Deal: energy saving for your home https://www.gov.uk/green-deal-energy-saving-measures/overview

UK Green Building Council (UKGBC), 2013. UKGBC Task Group Report: Retrofit Incentives https://www.gov.uk/green-deal-energy-saving-measures/overview

Visscher, H., Meijer, F., Macjen D., and Itard, L., 2016. Improved governance for energy efficiency in housing, Building Research & Information 44, 552-561

Zero Carbon Hub, 2014. Closing the gap between design and as-built performance http://www.zerocarbonhub.org/sites/default/files/resources/reports/Closing_the_Gap_Between_Design_and_As-Built_Performance-Evidence_Review_Report_0.pdf

Highlights

- Multiple EPC assessments carried out on dwellings to investigate consistency
- Analysed in context of the Green Deal but with wider relevance to EPCs
- Despite standardised methodology, lack of uniformity of outputs
- Implications on EPC-type approaches for quantifying energy efficiency discussed
- Combination of assessment and assessor gives rise to inconsistency