First evidence for the reliability of building co-heating tests

Richard Jack, Dennis Loveday, David Allinson & Kevin Lomas

Published online: 27 Mar 2017.

Submit your article to this journal ⧉

View related articles ⧉

View Crossmark data ⧉

RESEARCH PAPER

First evidence for the reliability of building co-heating tests

Richard Jack ⬤, Dennis Loveday ⬤, David Allinson ⬤ and Kevin Lomas ⬤

Building Energy Research Group, School of Civil & Building Engineering, Loughborough University, Loughborough, UK

ABSTRACT

This paper provides powerful evidence empirically demonstrating for the first time the reliability of the co-heating test. The test is widely used throughout Europe to measure the total heat transfer through the fabric of buildings and to calculate the heat-transfer coefficient (HTC; units W/K). A reliable test is essential to address the 'performance gap', where in-use energy performance is consistently, and often substantially, poorer than predicted. The co-heating test could meet this need, but its reliability requires confirmation. Seven teams independently conducted co-heating tests on the same detached house near Watford, UK. Despite differences in the weather and in the experimental and analytical approaches, the teams' final reported HTC measurements were within ±10% of the mean. With further standardization it is likely to be possible to improve upon this reproducibility. Furthermore, uncertainty analysis based upon a 95% confidence interval resulted in an estimated uncertainty in HTC measurements of ±8%. This research addresses persistent doubts about the reliability of the co-heating test. Avenues to further improvement of the test are discussed. This work helps to enable the test's wider adoption as a component of the regulatory process and thus improvements to standards of house construction.

KEYWORDS
co-heating; energy performance; heat transfer; heat-transfer coefficient; houses; *in-situ*; performance gap; performance assessment

Introduction

For regulatory purposes, the expected energy demand and carbon emissions of buildings are based on calculation tools, *e.g.* in the UK the Standard Assessment Procedure (SAP) (BRE, 2011, 2014). Such tools utilize the design drawings and building surveys to calculate the thermal performance of the building fabric based on construction properties as measured in laboratory conditions, and by a limited number of *in-situ* tests. However, whole-house performance measurements have provided considerable empirical evidence for the existence of a 'performance gap' between the predicted energy performance of a house and that actually measured *in-situ* (Gupta, Gregg, Passmore, & Stevens, 2015; Johnston, Farmer, Brooke-Peat, & Miles-Shenton, 2016; Palmer, Godoy-Shimizu, Tillson, & Mawditt, 2016; Stafford, Bell, & Gorse, 2012; Zero Carbon Hub & NHBC Foundation, 2010).

Evidence shows that, in the vast majority of cases, both the energy consumption and the total heat loss through the building fabric are higher (*i.e.* worse) than predicted. For instance, 30 of the 34 tests reported by Stafford et al. (2012) found measured fabric heat losses greater than predicted, with differences ranging from 1% to over 120%. In the UK, the housing stock accounts for about 30% of total energy consumption (DECC, 2011) with space heating accounting for 64% of this (DECC, 2012). In this context, it is clear that the performance gap has a substantial impact on national energy consumption. The presence of the performance gap has been demonstrated consistently in existing, newly built and recently retrofitted dwellings, including projects where energy performance was highlighted as an important factor (Gupta et al., 2015; Johnston et al., 2016). Given the existence of this performance gap, it is unlikely that further tightening of regulations based upon as-designed targets will have the desired effect (Visscher, Meijer, Majcen, & Itard, 2016).

The term 'performance gap' has been used flexibly, indeed loosely, to refer to various differences between what is expected and what actually occurs (de Wilde, 2014). Throughout this paper, the term refers specifically to a difference between the predicted and actual *thermal performance* of a dwelling, and not between the predicted and actual *energy use*.

A performance gap is likely to be caused by a combination of factors, including underperformance of individual building elements and a lack of airtightness, but also harder-to-detect thermal bridging at joints between materials and small areas of missing insulation. The thermal performance of the whole building is defined by this multitude of elements that would be practically impossible to measure individually; a test of the whole house is necessary to provide an accurate reflection of real performance.

The thermal performance of a building fabric can be described by two main heat-flow mechanisms: heat transfer through the fabric of the building, including the net energy flow through glazing and heat bridges; and heat lost due to air infiltration. In many countries, air tightness testing is routinely undertaken. For example, the 2006 UK Building Regulations (HM Government, 2006) introduced mandatory air-tightness testing for a sample of all newly built houses. The available evidence shows a consequential improvement in as-built airtightness (Zero Carbon Hub & NHBC Foundation, 2010), indicating that *in-situ* testing can diagnose faults, allowing them to be corrected, and act as an effective regulatory tool.

A similar regulatory and diagnostic tool would be useful for measuring the total heat transfer through the building fabric in both new builds and refurbishment. Such a tool could provide: a compliance check for new construction; quality assurance through measurements made 'before' and 'after' an energy-efficiency improvement; and a quality-control mechanism to support financial protection for both installers and occupants. To be accepted by the construction industry, the fabric heat-transfer measurements must be quick to undertake and reliable.

Tests to measure the fabric heat transfer of whole buildings can be broken down into two types: dynamic and quasi-steady-state. In dynamic testing methods, measurements of temperature and energy use are taken during periods of heating and cooling, carried out over a relatively short period of time (two to eight hours, much shorter than the several days required for quasi-steady-state tests). The measurements are then used in a simple lumped-parameter heat-transfer model to infer both the thermal mass and the heat transfer of a building. Examples of such tests are the primary and secondary terms analysis and re-normalization (PSTAR) and quick *U*-value of buildings (QUB) methods (Mangematin, Pandraud, & Roux, 2012; Mangematin, Pandraud, Gilles, & Roux, 2012; Palmer, Pane, Bell, & Wingfield, 2011; Subbarao, 1988). In the quasi-steady-state tests, the amount of energy required to maintain a constant, raised, indoor temperature is measured,

and the total heat-transfer rate is inferred by a simple energy balance. The most prevalent example of this method is the co-heating test, which has been in existence since the early 1980s (Everett, 1985; Siviour, 1981), but has been more widely used in the last 10 years in both the UK (Alexander & Jenkins, 2015; Guerra-Santin, Tweed, Jenkins, & Jiang, 2013; Jack, 2015; Jack, Loveday, Allinson, & Lomas, 2015; Johnston et al., 2016; Lowe, Wingfield, Bell, & Bell, 2007; Stafford et al., 2014; Stamp, 2015; Stamp, Lowe, & Altamirano-Medina, 2013; White, 2014) and the rest of Europe (Bauwens & Roels, 2014; Bauwens, Standaert, Decluve, & Roels, 2012; Meulenaer, Veken, Verbeek, & Hens, 2005). The co-heating test is the most commonly and widely used test of whole-building thermal performance at the time of writing.

The thermal performance of whole buildings is most often quantified by the heat-transfer coefficient (HTC). 'HTC' is interchangeable with a second term, the heat-loss coefficient (HLC), which has often been used when reporting co-heating results. 'HTC' has now been adopted as a standard term in line with the naming convention used in ISO 13790:2008, the international standard method for calculating space heating and cooling (BSI, 2008a; IEA, 2016). The HTC is a useful metric that describes the total, time-averaged, rate of heat transfer (in watts) from a building per-degree-Kelvin difference between indoor and outdoor air temperatures. Each building can be assumed to have a constant HTC, a value that is calculated as a metric in building energy models such as SAP. By measuring the HTC, the thermal performance of the whole building, as built, can be directly compared with the calculated performance, independent of occupant behaviour and weather conditions.

A method statement describing how to carry out the co-heating test was published by Leeds Beckett University in 2010 (Wingfield, Johnston, Miles-Shenton, & Bell, 2010), and was more recently updated to include further experimental guidance and, for the first time, detail on the data-analysis method (Johnston, Miles-Shenton, Wingfield, Farmer, & Bell, 2012). These method statements provide the basis for most co-heating testing conducted today because an industry-standard protocol for carrying out the measurements and for analysing the data is yet to be established, even though the co-heating test is already offered commercially (BSRIA, 2011; Stroma, 2015). The method published by Leeds Beckett was officially adopted by the UK Technology Strategy Board (TSB, now renamed as Innovate UK) as part of its building performance evaluation programme (TSB, 2012). Two particular issues with the co-heating test remain: (1) the length and invasiveness of the test,

which requires a house to be vacated for a period of at least two weeks and can only be carried out during the winter months; and (2) questions about the reliability of the test's results. These two issues have limited the wider use of the co-heating test by house builders and others, despite its proven value in the research field.

Against this background, the National House-Building Council (NHBC) Co-Heating Test Research Project (Butler & Dengel, 2013) investigated the reliability and practicality of the co-heating test. The project involved seven teams from different organizations successively undertaking co-heating tests on the same house as well as continuous measurements in an adjacent, nominally identical, dwelling. Some teams deliberately used a variety of testing and analysis approaches in order to explore the effect of the calculated HTC; but all teams declared a final reported value.

In November 2013, NHBC report NF54, *Review of Co-heating Test Methodologies*, was produced, which contained a background to co-heating testing and an initial analysis of the results (Butler & Dengel, 2013). It concluded that variable weather conditions, in particular heat gain resulting from solar irradiance, were the largest cause of uncertainty in the test results.

The objective of this study is to use this unique dataset to estimate the robustness, reproducibility and uncertainty of the currently available co-heating test in order to address persisting doubts about the reliability of tests that measure whole dwelling performance. This has relevance beyond the co-heating test itself by providing a way to measure trusted benchmark values against which the results of new, quicker or less invasive methods can be compared.

This paper describes the co-heating trials; compares the co-heating protocols used by the participating organizations; quantifies the reproducibility of the HTC calculated from each trial; estimates the uncertainty caused by the different methodologies used by each organization; and provides recommendations towards a best practice co-heating protocol.
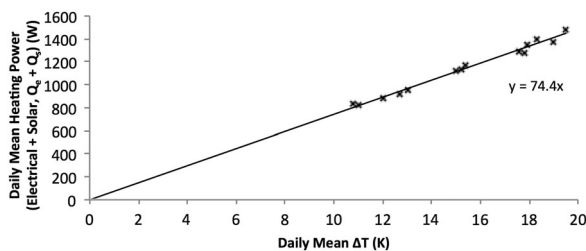

Figure 1. Example of a co-heating analysis plot. The heat-transfer coefficient (HTC) is taken as the gradient of a line of best fit, forced through the origin (in this example, HTC is 74.4 W/K).

Whilst further research will be required to understand the effect of testing in different buildings and locations, this research substantially advances the aim of producing a robust and reliable industry-agreed procedure.

Co-heating test

Basic procedure

The co-heating test uses a steady-state energy balance to calculate the total (both fabric and infiltration) heat transfer rate of a building including thermal bridging, with the result most commonly reported as a heat-transfer coefficient (HTC) with units of watts per Kelvin (see equations 1–3). The total heat input rate to the building, provided by electrical and solar heating, is deemed to be equal to the total heat loss rate from the building (Siviour, 1981). The energy balance is typically carried out using measurements that are averaged over a 24-h period:

$$\text{Electrical Heating} + \text{Solar Heating}$$
$$= \text{Fabric Heat Loss} + \text{Infiltration Heat Loss} \quad (1)$$

or:

$$Q_e + Q_s = Q_f + Q_i = (\Sigma U.A + 0.33N.V)\Delta T \quad (2)$$

$$\text{HTC} = (Q_f + Q_i)/\Delta T = \Sigma U.A + 0.33N.V \quad (3)$$

where Q_e is the rate of electrical heat input to the building (W); Q_s is the rate of solar heat input to the building (W); Q_f is the fabric heat loss rate from the building (W); Q_i is the infiltration heat loss rate from the building (W); HTC is the total HTC of the building (fabric plus infiltration) (W/K); $\sum U.A.\Delta T$ is the fabric heat-loss term, in which U is the U-value for each building element (W/m²K); A is the total area of each building element (m²); ΔT is the air temperature difference between the inside and outside of the building, referred to as 'Delta T' (K); $0.33N.V.\Delta T$ is the air infiltration heat-loss term, in which 0.33 approximates to the density of air multiplied by its specific heat capacity at 25°C (kJ K⁻¹ m⁻³); N is the air leakage rate in air changes per hour (1/h); and V is the internal heated volume of the building (m³).

During a co-heating test, the internal air temperature inside the building is maintained (using electrical heating equipment) at a constant raised level, commonly 25°C (Johnston et al., 2012), for one to three weeks. In order to ensure an even temperature distribution within the building, fans are used to mix the contained air, and all internal doors are held open. The rate of electrical energy consumption (Q_e) required to maintain this elevated temperature is recorded together with the weather

conditions, in particular the ambient air temperature and the solar irradiance.

To calculate the HTC, the solar gain (Q_s) must be included in the energy balance. The methods most commonly used to calculate the solar gains, as well as additional methods used by the organizations participating in the NHBC trials, are described below. The HTC can then be determined by plotting the rate of heat input (both electrical and solar) against the air-temperature difference between the inside and outside of the building. The gradient of a line of best fit, which is forced through the origin, is equal to the HTC (Figure 1). Each data point in Figure 1 represents the mean of measurements taken over a 24-h period.

Plotting daily averages assumes that any solar gains only have an impact on the internal thermal conditions on that day. In fact, of course, the impact can be carried over to subsequent days by storage in a building's thermal mass. The start and end of the daily averaging period can be chosen to try to minimize such carryover, but there is no commonly recognized definition of the start and end point of a day.

To separate the heat transfer through the fabric of the building (Q_f) from that due to infiltration (Q_i), the air leakage rate (N) must be measured. Usually a blower door test is used, but a tracer gas decay test is an alternative (BSI, 2001; Roulet & Foradini, 2002). Infiltration rate measurements could be carried out before a co-heating test, after a co-heating test, or both, and the results averaged. The reason for doing both is that the co-heating test may be causing additional cracking or drying out of materials, thereby altering the infiltration rate. The measured air leakage rate (N) and internal heated volume (V) can be substituted into the air infiltration heat-loss term ($0.33N.V.\Delta T$) in equation (3) to calculate the infiltration heat loss per degree Kelvin temperature difference between inside and outside (ΔT).
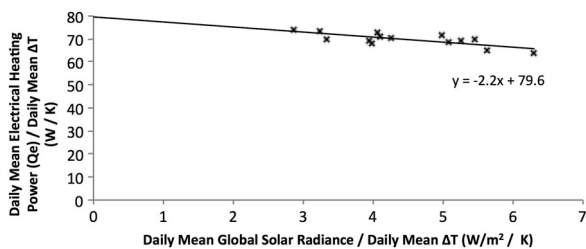

Figure 2. Example of the Siviour co-heating analysis method. The *y*-intercept is the HTC and the gradient is the solar aperture (in this example, HTC and solar aperture are 79.6 W/K and 2.2 m², respectively).

Accounting for solar gains

Even during the winter months, it is likely that significant solar gains will occur during co-heating tests, and accurately quantifying these is therefore an important part of the test. Six different methods were used in this project: Siviour analysis; Siviour plus regression; multiple regression; window transmission modelling; direct measurement; and using night or early morning measurements only. The first four methods require measurement of the total solar irradiance, the fifth method uses vertical solar-irradiance measurements on each facade, whilst the last method avoids the need for any solar-irradiance measurement. In co-heating tests, solar irradiance has most commonly been measured either horizontally or with a single south-facing vertical measurement (Stamp, 2015). Horizontal solar-irradiance measurement is most common in weather stations and does not suffer a directional or temporal bias, as is the case with vertical measurements. However, a horizontal measurement does not explicitly measure the irradiance falling on vertical facades through which the majority of solar gains are likely to occur (Stamp, 2015). The Leeds Beckett method recommends a vertical, south-facing, measurement of total solar irradiance (Johnston et al., 2012).

Siviour analysis

During the early development of co-heating analysis, Siviour (1981) proposed a graphical method to account for solar gains (referred to here as the 'Siviour' method). The daily mean electrical heating power is plotted against the daily mean global solar irradiance on a horizontal surface, with both terms divided by the temperature difference (Figure 2). The *y*-axis intercept of a linear trend line is then the HTC of the building, and the gradient is termed the 'solar aperture', the latter being a term that represents the equivalent area of glazing (in the same orientation as the solar-irradiance measurement, *i.e.* horizontal when using global solar irradiance) through which the solar gains have occurred (Everett, 1985).

Siviour plus regression

In an alternative approach, a Siviour plot is used to define the solar aperture of the building, which is then multiplied by the same daily mean solar irradiance measured in the same orientation as that used to define the solar aperture, to calculate an average solar heat input rate, in watts, for each day. This is then added to the average electrical heating power for the day and a linear regression is carried out with internal–external

temperature difference, as in Figure 1. This method will be referred to here as 'Siviour plus regression'.

Multiple regression

A third analysis method has been used by researchers at Leeds Beckett (Johnston et al., 2012). In this method, a multiple linear regression analysis is carried out using daily averaged data where electrical power is the dependent variable, internal–external air temperature difference and mean global solar irradiance are the independent variables, and the linear regression constant is assumed to be zero. The output of this analysis gives a correlation coefficient between electrical power and global solar irradiance, referred to as the solar aperture (as in the Siviour method). This value is multiplied by the mean global irradiance for each day to calculate the daily solar gain, which is then added to the daily mean electrical heating input, as in equation (1). The HTC is then determined using a co-heating plot, as shown in Figure 1; this method is described here as 'multiple regression'.

Two different methods for directly estimating the solar gains through the windows have also been employed. These methods assume that solar gains through opaque elements can be ignored.

Window transmission modelling

The 'simple window model' method uses the measured hourly global horizontal solar irradiance together with some form of modelling to estimate the irradiance falling on each glazed facade. The known window areas together with their G-values (a measure of total solar energy transmittance – the proportion of incoming solar energy transmitted into the building) are then used to calculate the solar gain. There are a range of models that could be used: simple models such as the approach set out in CIBSE guide A, Appendix 5.A10 (CIBSE, 2015), or dynamic models. Simple methods may not be able to account fully for site shading, hourly changes in solar irradiance or the external inter-reflection of solar irradiance. No account was taken of shading from trees or other buildings when these methods were used in this project.

Direct measurement

To avoid the need to translate global irradiance to vertical irradiance, the solar gain to each facade can be directly measured using pyranometers mounted on each glazed facade. Simple equations can then be used to estimate the solar irradiance entering the space through each window. One such equation is equation (4), which is based upon equation (6) taken from the 2012 Standard Assessment Procedure (SAP) (BRE, 2014):

$$Q_s = \Sigma_i T_r \times A_{wi} \times S_i x g_i \tag{4}$$

where Q_s is the total daily mean solar gain (W); T_r is the ratio of typical average transmittance to that at normal incidence; A_{wi} is glazed area (excluding window and door frames) on facade i (m^2); S_i is the daily mean total solar irradiance (as measured by a pyranometer) on facade i (W/m^2); and g_i is the total solar energy transmittance factor of the glazing on facade i at normal incidence.

Night measurement

Analysis using measurements taken during the night or early morning only removes the uncertainties and complexities surrounding the calculation of solar gains. Different night or early morning periods can be chosen. The data points on the regression plot (Figure 1) are therefore averaged over a much shorter time period than that of a whole day. Thus, the position of the data points is compromised, although the uncertainty over Q_s is removed. No account is taken of any heat stored in thermal mass during the day.

Accounting for wind speed

Variations in wind speed affect the infiltration rate and the heat transfer at external surfaces, although in well-insulated buildings the second of these is of minor importance. Wind speed measurements are often taken on-site, though it can be difficult to find a suitable location that is free of local disruptions to the airflow, such as other buildings or trees (which cause an inaccurate measurement to be recorded). Wind measurements obtained from a local weather station can also be used if suitable data exist.

Multiple regression analysis can be used to account for the influence of variations in wind speed (Johnston et al., 2012). Wind speed is simply included as an additional independent variable, together with solar irradiance and indoor-outdoor temperature difference. The HTC can then be determined directly from this multiple regression analysis, or the regression coefficient associated with wind speed is multiplied by the daily mean wind speed to calculate the effective increase or decrease in heat transfer due to the wind speed on each day. In this way, the daily mean heating power is adjusted to take account of the wind speed that day, and the HTC is effectively determined at zero wind speed. Reporting the HTC for a wind speed of zero aligns with the approach taken for solar irradiance. Johnston et al. (2012) note, however, that in their experience this

approach is inherently prone to inaccuracy due to complex interrelations between variables such as sheltering, orientation, airtightness, location and leakage paths specific to each building.

Testing programme

The Co-Heating Test Research Project described here was set up by the NHBC Foundation in cooperation with the Building Research Establishment (BRE). The BRE is an independent research organization focused on all aspects of buildings with extensive experience of energy performance testing. The BRE provided access to a pair of identical adjacent detached test houses for the project, denoted as 50.3 and 50.4. These were located at the BRE site at Garston, just north of London, UK (latitude 51.7, longitude –0.4). The houses are of a simple rectangular plan form over two floors (ground and first floors). They were located in an open area with little overshading or wind sheltering, with the front facade facing south (Figures 3 and 4).

The purpose-built test houses were constructed in 1995 to contemporary Swedish building standards. At the time of testing, the walls were understood to consist of an outer leaf of bricks, followed by a 50 mm air gap, 13 mm of fibreboard, 170 mm of Rockwool insulation, 9 mm of plywood, 45 mm of Rockwool insulation and finally 13 mm of plasterboard on the internal surface. The houses were triple-glazed, with 240 mm of loft insulation installed, and with 220 mm of insulation installed beneath the suspended ground floor. Both houses had an unheated attic space. This resulted in a thermal performance similar to, but slightly worse than, that required by 2013 UK Building Regulations (HM Government, 2013) (Table 1). Purpose-made ventilation paths were temporarily blocked during all tests.

The houses had a reported HTC calculated by SAP of 68.4 W/K (65.9 W/K for the fabric and 2.5 W/K for the infiltration) (Butler & Dengel, 2013). This result was

Figure 3. South-facing front facades of test houses 50.3 (left) and 50.4 (right).

made available to all participants prior to the start of testing.

Seven teams from different organizations participated in the study with each organization employing its own approach to conducting a co-heating test. Tests were carried out one at a time, in serial fashion, between December 2011 and May 2012. The testing teams are referred to in this paper by the randomly assigned letters A–G.

Over the 2011–12 winter heating season, each of the teams was allocated approximately two weeks in which to perform a co-heating test in building 50.4 (Table 2). No pre-heating of the house was carried out before any tests, though in most cases the testing periods almost immediately followed each other, with little time for the building to cool between tests.

Each organization was required to independently report their measured HTC together with a summary of the testing and data-analysis methods used. A prescriptive list of results to be reported was not specified. No guidance was given to any team on the design and planning of their data-collection or analysis procedures, this being a deliberate decision in order to enable the full range of possible co-heating methods to be used. It is possible that this will have led to a larger variation in the results than if a specific method had been prescribed.

In parallel, the BRE carried out a separate co-heating test continuously in the adjacent house (50.3) covering the entire period between December 2011 and September 2012. The test was carried out according to the method described in the subsequent report NHBC NF54 (Butler & Dengel, 2013). An internal set-point temperature of 25°C was used initially, which was later increased to 30°C (on 22 May 2012) in order to maintain a higher internal–external air temperature difference during the warmer summer months.

Testing protocols

The seven teams used an experimental method broadly similar to that published by Leeds Beckett University (Johnston et al., 2012), but with variations in the detail (Table 3). All teams used electric heaters with electric fans to circulate the warmed air, with a fan placed adjacent to each heater. This combination of heater and air mixing fan is referred here to as a 'heating and mixing station'. Each team distributed heaters and fans around the house according to their own regime, with the majority placing one in each large room of the house to the extent that their equipment availability would allow. All teams used different sensors for measuring the resulting space temperature from those used to control the heaters. Some teams also took the opportunity to
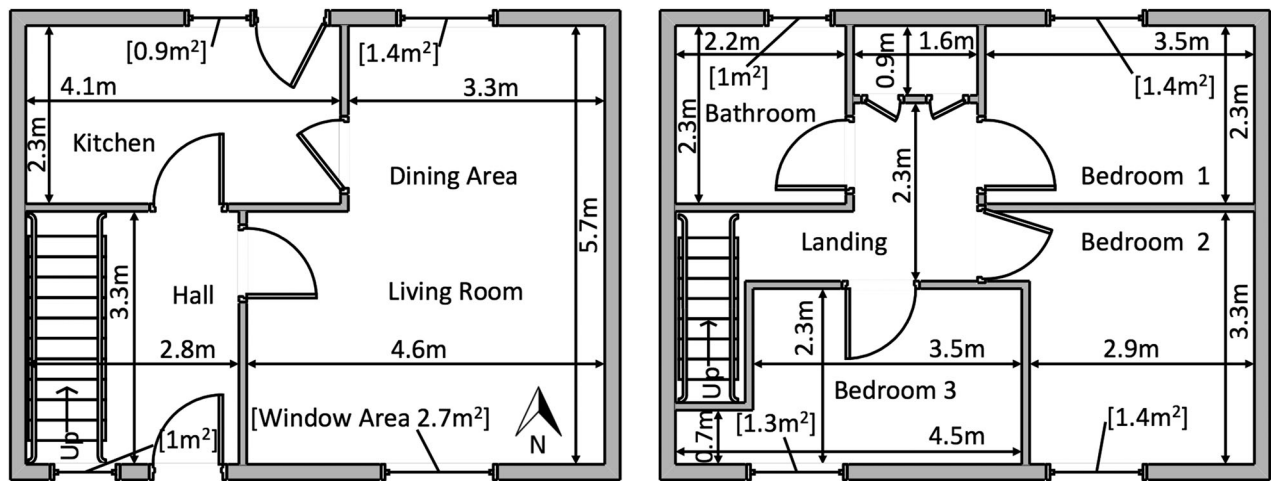
Figure 4. Floor plans of test houses 50.3 and 50.4 (they are identical), ground floor to the left and first floor to the right.

make additional complementary measurements, such as infrared thermography and U-value measurements using heat flux plates.

In this project, many teams carried out only one air-infiltration-rate measurement. This was because, unlike a newly built home, the BRE houses had already undergone a sustained period of co-heating, and indeed a long history of testing in general, and so air-tightness was very unlikely to be altered by their co-heating test. The infiltration rate as measured by blower door tests and tracer gas decay methods was very similar, though there was more variation in the results measured by the tracer gas-decay method, as evidenced by the higher standard deviation (Table 4). It should be noted that the infiltration rate under ambient conditions is affected by factors such as wind speed and internal–external pressure difference, therefore the large standard deviation of the measurements taken by the tracer-gas method could reflect actual changes in infiltration rate. Three teams, C, E and F, used infiltration rate measurements to report the HTC disaggregated into fabric and infiltration heat losses.

The additional heat flux and infrared thermography measurements made by some teams were not necessary for the calculation of the HTC. However, they could be helpful for diagnosing the causes of heat loss in the case of unexpectedly high measured HTC values. The raised indoor air temperatures used in the co-heating test can help improve the accuracy of U-value measurements and make sources of heat leakage more evident in infrared thermography surveys.

The margin of uncertainty in co-heating test results is associated not only with the accuracy of the equipment used but also with the data-analysis methods employed to translate measurements of irradiance and wind speed into effects on the HTC. Only two teams, C and D, reported their HTC result together with an uncertainty range. In each case, this was defined as one standard error in the average daily internal–external temperature versus heating power regression, to either side of the reported HTC. This could reflect the difficulty in accurately defining the measurement uncertainty of the test, particularly the uncertainty due to the solar gains.

Table 1. Comparison between design U-values for the test houses used in the National House-Building Council (NHBC) co-heating project and those stipulated in the 2013 edition of Part L of the Building Regulations (HM Government, 2013).

Building element	Design U-value (W/m²K)	2013 UK building regulations' limiting U-value (W/m²K)	2013 UK building regulations' notional specification U-value (W/m²K)[a]
External walls	0.21	0.30	0.18
Floor	0.21	0.25	0.13
Roof	0.16	0.20	0.13
Windows	1.85	2.00	1.40
External doors	1.00	2.00	1.00

Note: [a]In addition to a minimum elemental performance (column 3), UK building regulations specify a minimum performance level calculated for the whole dwelling that can be achieved with different combinations of elemental performance. The notional specification provides an example of a specification that reaches the minimum performance level for a whole dwelling.

Table 2. Testing periods and duration of testing for each of the seven participating teams.

Testing team	Dates of test	Test length (days)
A	22 December 2011–2 January 2012	12
B	24 January–6 February 2012	14
C	10–24 February 2012	15
D	28 February–15 March 2012	17
E	16 March–2 April 2012	18
F	3–16 April 2012	14
G	23 April–8 May 2012	16

Reproducibility of HTC measurements

Reported HTC values

There was a failure of energy metering equipment during the test carried out by team G, which meant that a full set of energy-consumption data was not collected for the majority of their testing period. Team G's results were therefore excluded from the further analysis reported in this paper. The HTC values reported by each of the other teams, following their independent testing on the same test house (BRE

Table 3. Comparison of key components of the testing and analysis protocols used by each team.

Testing method		A	B	C	D	E	F	G
					Testing organization			
Heaters and fans	Number	Not reported	5	5	5	3	8	6
	Convector	●						
	Fan		●	●	●	●	●	●
Temperature control	On/off thermostat		●		●	●		●
	Proportional-integral-derivative (PID)	●		●			●	
	Set point (°C)	25	25	25	24	27	25	26
Electricity use measurement	Meter	●			●		●	
	Plug		●	●	●	●	●	●
Temperature measurement locations	Every room	●	●	●	●		●	
	Heated rooms							
	Unheated rooms					●		●
Infiltration measurement	Blower door before	●	●				●	
	Blower door after	●	●			●	●	●
	Tracer gas decay		●[a]	●[a]	●[b]		●[a]	
Solar measurement	Horizontal	●			●	●		
	Vertical south		●	●	●		●	●
	Vertical each facade							●
Other measurements	Heat flux		●	●	●			
	Infrared thermography		●	●			●	●

Data-analysis methods		A	B	C	D	E	F	G
Daily averaging	Start and end hour	00:00–23:59	00:00–23:59	06:00–05:59	09:00–08:59	00:00–23:59	18:00–17:59	00:00–23:59
Accounting for solar gains	Siviour	○		○	●		○	○
	Siviour plus regression	●		●	●			
	Multiple regression		●	○		●	○	
	Window transmission modelling	○[c]						○[d]
	Direct measurement						●	
	Night measurement				○[e]			●[f]
Accounting for wind speed	Multiple regression	●	●		●	●	●	●
	None			●				
Uncertainty in calculated HTC	Reported				●	●		
	Not reported	●	●	●			●	●

Notes: 'Heat flux' refers to *U*-value measurements using heat flux plates.
●, Final HTC reported using this method; ○, alternative method, but not the one for which the HTC results were reported.
[a]Tracer gas decay measurements made daily.
[b]One-off tracer gas measurement.
[c]Calculated using measured global horizontal irradiance and CIBSE Guide A, Appendix 5.A10 method (CIBSE 2015).
[d]Calculated using Ecotect software.
[e]Measurements made from 06:00 to 08:00 only used.
[f]Measurements made from 22:00 to 04:00 only used.

Table 4. Mean infiltration rate as measured by blower door tests and tracer gas-decay methods and the standard deviation of these.

Test method	Mean air-change rate (1/h)	Standard deviation (1/h)
Blower door[a]	0.076	0.002
Tracer gas decay[b]	0.076	0.010

Notes: [a]Blower door results are converted to air-change rate at ambient pressures using the 1/20 rule of thumb (Sherman, 1987). This method was developed in the US and may not be appropriate for use in UK homes. Values from the three organizations that reported their blower door results result contribute to these data.
[b]Eighteen separate tracer gas measurements contribute to this result.

Table 5. Uncertainty in the HTC resulting from a 1°C change in daily average internal–external air temperature difference (ΔT).

Analysis method	HTC (W/K)	HTC uncertainty due to ΔT measurement (W/K)	
		ΔT −1°C	ΔT 1°C
Multiple regression	52.6	2.7 (5.0%)	−4.0 (−7.5%)
Siviour	57.7	2.8 (4.9%)	−2.6 (−4.6%)
Siviour plus regression	50.8	1.9 (3.7%)	−1.8 (−3.6%)
Window estimation	64.8	3.7 (5.7%)	−3.4 (−5.2%)

Note: Figures in parentheses show the uncertainty as a percentage of the original HTC.

Building 50.4, the right-hand house in Figure 3) over the period, December 2011 to May 2012, are shown in Figure 5.

For this dataset, the mean HTC is 65.8 W/K, with a standard deviation (for small datasets) of 3.2 W/K, a standard error of the mean of 1.3 W/K, and hence a 'population' mean of 65.8 ± 2.5 W/K at the 95% confidence level and 65.8 ± 3.7 W/K at the 99.75% confidence level. Despite variations in weather and in the teams' data-collection and analysis methods, the reported HTC values for all six teams fell comfortably within a ±10% range around the mean. The largest differences between a reported result and the mean in this dataset are −7.0% and 5.8% for teams E and F respectively. Four of the six HTC results fall within the 99.75% confidence interval.

Two teams, C and F, reported the portion of the HTC that was due to infiltration as 5 and 4.8 W/K respectively.

Effects of the different data-collection methods

Differences in the methods that teams used to carry out their co-heating testing as described above included: the way that temperature was regulated, measured and supplied; the method used to measure energy consumption; the location and extent of weather monitoring equipment; the measurement of the infiltration rate; and the other supporting measurements made.

Teams C, E and F reported the differences between the indoor air temperature measured in different rooms. Team C, using five heating and mixing stations, observed a slightly lower (approximately 1°C) temperature in rooms without a station in place. This could be pertinent to the approach employed by team G, which only installed temperature sensors in the rooms in which heaters were installed, potentially leading to an overestimation of the average internal temperature. Team F placed a heating and mixing station and temperature sensor in each room, and observed temperature variations of less than 0.5°C throughout the house. Team E used only three heating and mixing stations and measured a variation of approximately 2°C between the warmest and coolest rooms during the testing period. Whilst this may have been caused by the relatively small number of heaters and air movement fans used, it could also have been influenced by the placement of temperature sensors which were located at the perimeter of the five unheated rooms only. A further contributing factor could have been team C and F's use of proportional-integral-derivative (PID) temperature control compared with team E's on/off thermostatic temperature control (Table 3).
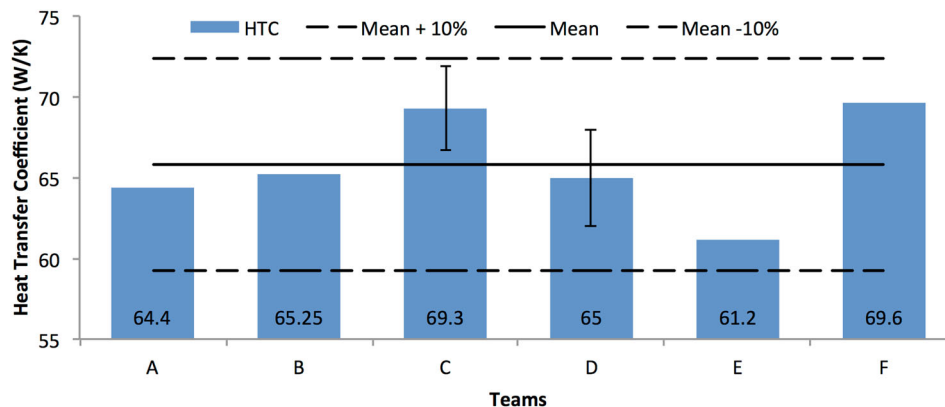

Figure 5. HTC reported by each team (excluding team G).

Comparison of HTC values and solar-gain calculation method

As shown in Table 3, several teams applied more than one solar gain analysis method before selecting one to calculate their final reported HTC value, effectively generating a larger sample of 13 measured HTC values.

It is not possible to state definitively the 'correct' value for HTC and therefore announce which of the analysis methods is the most accurate. Given the small number of applications of each method, neither is it possible to observe significant relationships between the data-analysis method employed and HTC value obtained. Despite the diverse analysis methods used, 10 of the 13 reported results fell within ±10% of the mean HTC of the whole sample: 64.8 W/K (Figure 6). The extreme values differed from the mean by −18% and +12%. The three results that fell outside of ±10% of the mean HTC for this dataset were reported by team A using the window-transmission modelling method (72.5 W/K) and team F, who were new to HTC testing, using the Siviour (56.7 W/K) and multiple regression (52.9 W/K) methods.

The results indicate that the co-heating test is reasonably robust to different measurement methods, weather conditions and data-analysis methods. If standardized testing and data-analysis protocols were defined and employed by trained and experiences teams, then better reproducibility than observed in these tests would probably be achieved.

HTC uncertainty analysis

An analysis was carried out to investigate the aggregate measurement uncertainties in three key measurements: internal–external temperature difference, electricity consumption and solar gains. The analysis allows the total uncertainty in the measurement of the HTC by co-heating tests to be determined and provides new insight into which elements of the co-heating test contribute the largest uncertainty to the measured HTC value. The uncertainty analysis used data collected by BRE in building 50.3 during the continuous co-heating test carried out between 12 February and 11 September 2012.

In this uncertainty analysis, it is assumed that a best-practice co-heating method (as identified in this paper) is applied, and properly calibrated sensors used. A best-practice co-heating method is considered to include temperature measurement in each room, the use of a heater and air mixing fan in each major room, electricity use measurement at the service meter, and solar-irradiance measurements taken in an unshaded location within reasonably close proximity (less than 20 km) of

the test building. The solar-irradiance measurements taken for this analysis were measured in a horizontal orientation.

Uncertainty in temperature measurement

Measurement of the internal–external air temperature difference is one of the key parts of the co-heating test. Temperature sensors such as thermistors offer accuracy of ±0.2°C, while the placement of sensors can introduce an added (systematic) uncertainty, as air temperature will vary within rooms. The internal–external air temperature difference is generated by two measurements, both prone to these uncertainties. Therefore, an uncertainty of ±1°C in the measurement of internal–external air temperature difference was assumed. This variation was chosen based upon temperature variations observed throughout the building in this project. It should be noted that there will, of course, be a considerably smaller variation in internal temperature during a co-heating test than there would be in most houses due to the careful temperature control applied.

The resulting uncertainty in the calculated HTC due to uncertainty in temperature measurement was estimated by adjusting the daily average internal–external temperature difference by 1°C for each day during the testing period, then calculating the HTC with this new dataset via four solar-gain analysis methods. The uncertainty is reported (Table 5) as the resulting change in HTC and, in parentheses, the percentage difference in the HTC compared with the value calculated using the original dataset.

Uncertainty in electricity consumption measurement

The same uncertainty analysis method was applied to the measurement of electrical power consumption (Table 6), with an uncertainty of ±5% in the electricity consumption chosen. This choice was based on regulations that

Table 6. Uncertainty in the HTC due to a ±5% change in the measured electrical power consumption.

Analysis method	HTC (W/K)	HTC uncertainty due to electricity consumption measurement (W/K)	
		−5%	**5%**
Multiple regression	52.6	−2.3 (−4.3%)	3.1 (5.8%)
Siviour	57.7	−2.9 (−5.0%)	2.9 (5.0%)
Siviour plus regression	50.8	−2.5 (−5.0%)	2.5 (5.0%)
Window estimation	64.8	−2.0 (−3.0%)	2.0 (3.0%)

Note: Figures in parentheses show the uncertainty as a percentage of the original HTC.

require an electricity service meter to have an accuracy of at least ±3.5% (HM Government, 1998). A further uncertainty of ±1.5% was added to account for the accuracy with which the service meter was monitored, based on comparison between measurements and service meter readings taken at the start and end of monitoring.

Tables 5 and 6 show that the uncertainty in measurement of temperature and power consumption result in an uncertainty of approximately ±5% in the calculated HTC, with a range of 3–7.5% depending on which data-analysis method is used.

Uncertainty in the calculation of solar gains

The uncertainty in the calculated HTC that results from the calculation of solar gain is more difficult to quantify as it comprises the uncertainty entailed in the measurement of irradiance (which is approximately ±5% for a first-class pyranometer, commonly used for research applications; ISO, 1990) plus that introduced by the assumptions inherent in each analysis method. As the results of this paper have demonstrated, accounting for solar gains is the largest area of variation in the co-heating protocols currently in use. Given the continuing ambiguity in how best to account for solar gains, it is important to quantify accurately the uncertainty introduced into the final HTC measurement by the estimation of solar gains.

To quantify the uncertainty associated with the selection of the data-analysis method, the variation in the HTC values calculated using each of the four methods for a number of two-week samples of the extended dataset was compared. The daily averaged data collected between 12 February and 30 April 2012 was separated into a series of samples, each of 14 days in length. This period includes all data collected within the heating season, defined as the start of October to the end of April (BRE, 2011), during which co-heating tests are usually carried out. The periods overlap in order to give the maximum possible number of samples. The sampling 'window' moves forward by a day at a time, so that the first sample is from 12 to 25 February, the second from 13 to 26 February, and so on. In total, this process results in 66 overlapping samples. This approach allows an assessment to be made over a range of weather conditions (Figure 7, lower graphs).

The uncertainty is thus calculated for a period during which UK buildings would be heated. During the summer, when solar irradiance is greater and day length longer, the solar heat gain in buildings is likely to be much higher. Solar gain would thus form a larger proportion of the total heat input and so any uncertainty in the solar-gain calculation method would have a

greater relative impact on the total uncertainty of the calculated HTC. Also, the solar conditions are more complex in summer as the sun moves through a much wider range of orientations and azimuths. (The limitation to winter-only data was not necessary for the measurements of temperature or electrical power consumption as the uncertainty associated with each is not affected by changes in weather conditions.)

There is naturally a variation in the inherent HTC at different times, which is independent of the solar gain analysis method used, e.g. because of moisture movement, changes in the infiltration and changes in the external surface heat transfer coefficients due to the wind etc. The drift in the average HTC calculated for each of the 66 samples is shown in Figure 7 (upper graph). Averaged across all 66 samples, the four analysis methods did, however, give very similar HTC values: 65.9–66.8 W/K (Table 7).

To allow for the inherent variation in HTC, for each sample the mean HTC calculated by all four analysis methods was subtracted from the HTC calculated by each individual method. This process returned a dataset with 264 points (66 samples and four analysis methods). It was then possible to determine, for each method, the 95 percentile positive and negative difference from that method's mean HTC (mean ± 1.96 SD), and also to express this difference as a percentage of that method's mean HTC (to enable direct comparison with the uncertainty in the measurement of temperature and electricity consumption).

It is clear from Table 7 that each method sometimes produced a higher HTC value than the mean and sometimes a lower value (there are both lower and upper differences) and that, for each method, the upper and lower 95 percentile intervals were of roughly similar magnitude. In other words, there was no obvious systematic over- or underestimation by any method. It is also clear, however, that the window analysis method was more variable than the others, indicating, perhaps,

Table 7. Variation in the HTC as a result of the solar gain analysis method.

Analysis method	Mean HTC (W/K)	95% Confidence interval	
		Lower (W/K)	Upper (W/K)
Regression	65.9	−2.0 (−3%)	1.5 (2%)
Siviour	66.2	−0.9 (−1%)	1.5 (2%)
Siviour plus regression	65.9	−1.2 (−2%)	1.1 (2%)
Window estimation	66.8	−2.9 (−4%)	4.0 (6%)
All methods	66.2	−2.0 (−3%)	2.3 (4%)

Note: Figures in parentheses show the variation as a percentage of the original HTC.
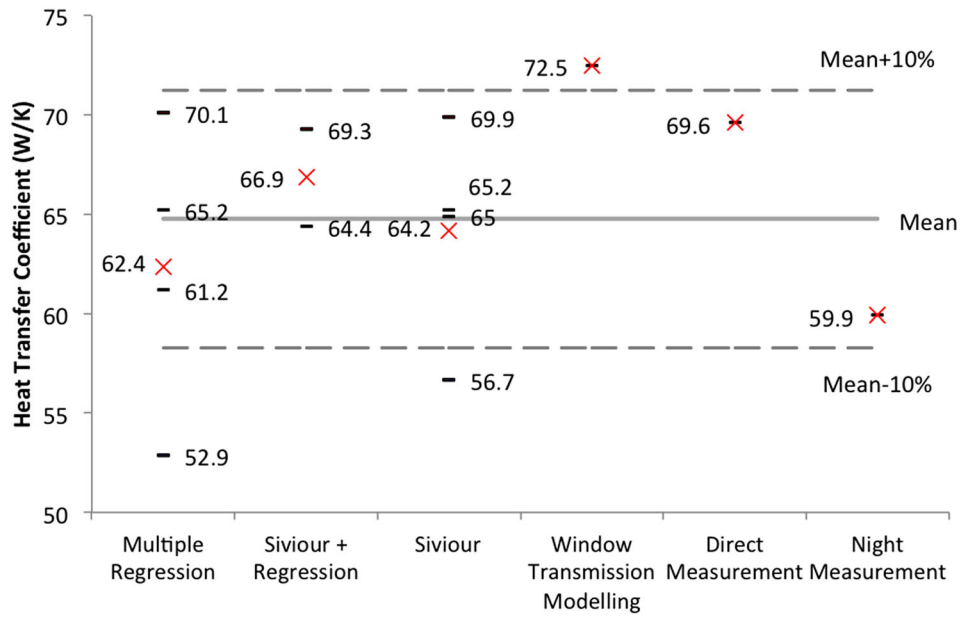
Figure 6. Comparison of calculated HTC values disaggregated by the solar-gain calculation method. The mean result for each analysis method is denoted by an 'X'. The mean HTC for the whole sample (including all solar-gain calculation methods), and ±10% of this value are also shown. See Table 3 for which team applied which method.

greater sensitivity to some feature of the ambient weather conditions.

Whether or not the variation, or indeed the mean HTC, calculated by any method is, or is not, a true reflection of actual physical phenomena cannot be determined on the strength of the data available here. Therefore, all results were given equal credibility and the overall 95 percentile confidence intervals calculated by pooling all 264 differences (Figure 8) was taken as the uncertainty in the measured HTC introduced by the solar gains

analysis method; this produced values of –2 W/K (–3%) to 2.3 W/K (4%) (Table 7).

In addition to the uncertainty contributed by the analysis of solar gains, uncertainty is introduced by the solar-irradiance measurement itself. The effect of this uncertainty on the HTC was calculated using the same uncertainty analysis that was adopted for temperature and electricity consumption measurement. An uncertainty of ±5% in the solar-irradiance measurement was chosen based on the assumption that a first-class pyranometer
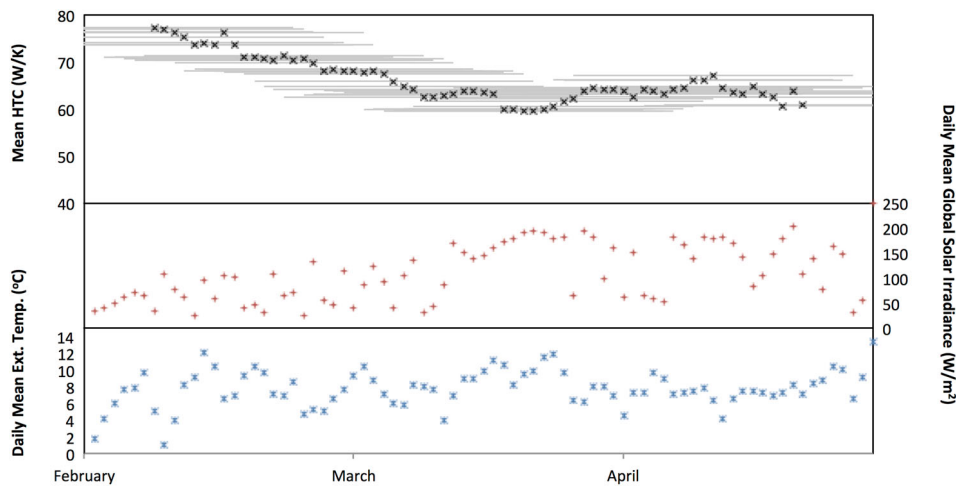

Figure 7. (top to bottom) Mean HTC (across all four analysis methods) for each 14-day sample (the midpoint of each sample is marked by a cross and the whole sample by a horizontal line), daily mean global horizontal solar irradiance and daily mean external temperature during the winter sample of the extended co-heating test.

is used (ISO, 1990). Uncertainty analysis was carried out using the window-estimation method, which was the method with the most variation (see above), and resulted in an uncertainty of ±2% in the calculated HTC.

Combined uncertainty in the HTC

Overall, the uncertainty in the calculated HTC due to the uncertainty in the internal–external air temperature difference and the electrical power consumption are each about ±5% on average across all data-analysis methods. A further ±2% is introduced from the measurement of solar irradiance and there is an additional uncertainty that varies depending on the solar-analysis method used. An approximation of the total uncertainty was calculated from the quadrature sum of the influence of each (Lomas & Eppel, 1992). This produced similar 95 percentile uncertainties for all four analysis methods (Table 8). The regression method has the largest uncertainty range, -9% to 8%, whilst the Siviour plus regression method had the smallest, ±6%. However, based on this limited study alone, it is difficult to conclude that one method has greater reliability than another.

The mean calculated uncertainty across all analysis methods, and hence the estimated general uncertainty in the HTC measurement by a co-heating test, was ±8% (Table 8). This estimated total uncertainty is similar to the actual variability in the measured HTC values for house 50.4, which was ±10% (Figure 5). The slightly lower uncertainty resulting from this analysis could be expected given that all the data from house 50.3 were collected using the same method and with the same equipment.

Discussion

The structure of the reported project has allowed a unique insight to be made into the co-heating test, and in particular into the range of approaches that can be

used to carry out ostensibly the same test. Given the breadth of measurement equipment, temperature control schemes and data-analysis methods, as well as differences in the weather conditions experienced during the periods of testing, it is noteworthy that the spread in the final reported HTC values was just ±10%. Several teams also gave values calculated by alternative data-analysis methods, and the results of all calculations were within –18%, 12% of the mean of these results. The uncertainty analysis produced an estimate of the uncertainty in HTC values derived from co-heating tests of ±8%. The uncertainty in co-heating results determined by this study is very similar to those reported by Alexander and Jenkins (2015) and Stamp (2015), both articles reporting measurement uncertainties of ±10% or less, given suitable testing conditions. By comparison, observed differences between the as-built HTC and the predicted HTC can be 100% or more (Johnston et al., 2016; Stafford et al., 2012). That is, the fabric heat loss from the actual constructed building can be twice the predicted value. A co-heating test protocol that is accurate to within 8–18% can, therefore, make a very valuable contribution to reducing this 'performance gap'. The co-heating test could be applied routinely as a quality assurance tool, but more importantly, in the fullness of time, it could become a regulatory requirement for both domestic new-build and refurbishment projects. There are, however, some practical questions still to be answered, and some are discussed here.

The range in the results presented here may appear lower than that reported in NHBC report NF54, in which it was stated that 'the maximum uncertainty in the results from the co-heating tests and the SAP equivalent was 17%' (Butler & Dengel, 2013, p. ix). This is because different methods of reporting the range in the results have been used in the NHBC report and in this paper. This paper has focused on the final HTC values reported by each team, and each measured result was compared with the mean of all measured results, while in the NHBC report the HTC calculated by SAP (BRE,

Table 8. Results of the uncertainty analysis showing the measurement uncertainty according to each data-analysis method and the overall uncertainty.

| Analysis method | HTC uncertainty due to measurement | | | | | | | | Total 95% confidence interval | |
| | ΔT | | Electrical power consumption | | Solar analysis method | | Solar irradiance | | | |
	–	+	–	+	–	+	–	+	–	+
Regression	5%	–8%	–4%	6%	–3%	2%	–2%	2%	9%	8%
Siviour	5%	–5%	–5%	5%	–1%	2%	–2%	2%	7%	7%
Siviour plus regression	4%	–4%	–5%	5%	–2%	2%	–2%	2%	6%	6%
Window estimation	6%	–5%	–3%	3%	–4%	6%	–2%	2%	7%	9%
Across all methods	5%	–5%	–4%	5%	–3%	4%	–2%	2%	8%	8%

2011) was used as the benchmark for comparison. The benchmarking of HTC measurements based on the predictions of a model was not used in this paper because models are simplifications of reality, and are known to produce predictions of energy demand that can be very different from those observed in practice. Indeed, measurement is needed precisely because HTC values calculated by models such as SAP are invariably lower (*i.e.* predict better thermal performance) than those observed *in-situ*.

Several key points can be made from the comparison of data-collection methods. Maintaining a constant internal air temperature throughout the building is vital in co-heating to ensure accurate measurements of the whole building's performance. It is clear that this is best achieved through the use of many heaters and mixing fans, with at least one in each large room. This supports the recommendations of several previous works (Johnston et al., 2012; Mangematin, Pandraud, & Roux, 2012; Stamp, 2015). The methodological comparison also suggests that a PID temperature control regime is beneficial in further ensuring that a constant temperature is achieved.

The issues caused by sensor failure in team G's test highlights a practical difficulty in performing co-heating tests remotely, which is a common practice due to access restrictions and the location of test sites. The value of a back-up measurement system is clear and is especially pertinent given the time-consuming and invasive nature of the co-heating test. Remote monitoring would further alleviate this problem in wider practice by allowing the tester to identify any failures immediately.

The use of in-line plug meters for every heating and mixing station has the advantage of allowing a disaggregation of the heat input to each space. Team F's analysis showed a higher heat input to the kitchen and bathroom, which coincided with leakage paths identified by an infrared thermography survey.

Differences in the data-collection methods observed in this project were minor in comparison with those observed in the data analysis, with a best-practice approach clearly emerging. The findings from the comparison of the data-collection methods presented in this paper, alongside the Leeds Beckett University method statement (Johnston, et al., 2012), provide a clear and complete set of guidelines that will lead to consistent and repeatable data collection.

The method used to account for solar gains remains an area of ambiguity and was the area of most variation between teams. The results of the comparative analysis (Figure 6) indicate that the Siviour plus regression method may offer a particularly repeatable result, and was shown to have a marginally lower response to uncertainties in temperature and energy consumption measurement and the estimation of solar gains (Table 8). Given the small sample size, no conclusive recommendations as to the most appropriate data-analysis method can be given, though the uncertainty analysis showed that in this case the chosen data-analysis method had little impact on the result, with all four methods resulting in very similar HTC values. Encouragingly, the results of both the uncertainty and comparative analyses suggest that the co-heating test is robust to small changes in the solar gain analysis method, and therefore that any of the methods studied would be suitable for general use. Despite this, the variety of analysis methods used to determine solar gain necessitates that the particular method chosen in each case is an essential part of the reporting of co-heating test measurements so as to ensure transparency in the result. Each method requires similar measurements to be undertaken, except the 'direct measurement' method, which requires additional
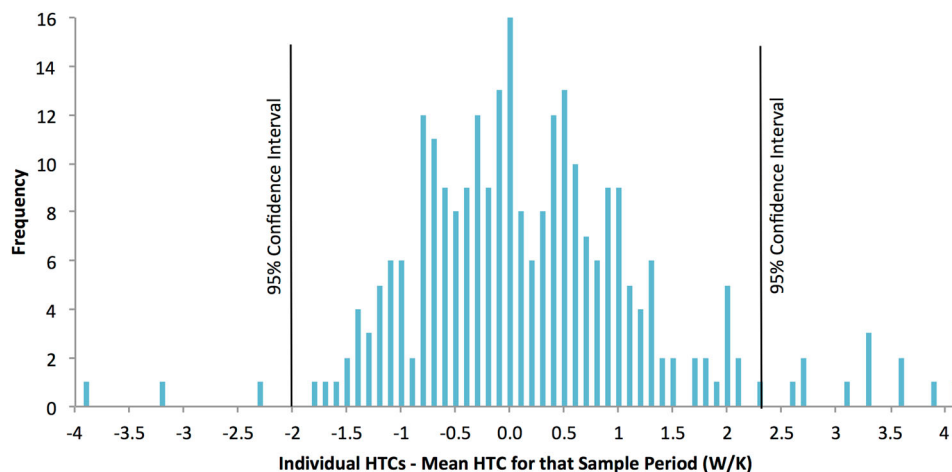

Figure 8. Frequency distribution of all 264 values of HTC differences produced by the four different solar gain analysis methods.

measurements of solar irradiance to be taken on each facade. Due to the additional measurements required, the 'direct measurement' method has a slightly more complex data-gathering protocol and higher equipment costs. Any definitive co-heating test protocol will, of course, need to define a single analysis method.

Several teams chose to report the HTC calculated using alternative methods to account for solar gains, in addition to reporting their final HTC by their preferred method. There was a larger range of –18%, 12% from the mean HTC, in this extended sample of HTC calculations. This larger range was caused by two outlying results, both reported by team F. If team F's results are removed from this sample, then the remaining 10 results all fall within ±10% of the new sample mean: 66.3 W/K. Team F carried out their test in April (Table 2); it was the last to be carried out and was towards the end of what has traditionally been considered the winter co-heating season in the northern hemisphere. Team F declared a final reported HTC value of 69.6 W/K, which was calculated using the 'direct measurement' method; this falls within ±10% of the mean HTC. This may indicate that this analysis method is more robust than others for use in periods of high solar irradiance, but the sample size is much too small to draw any reliable conclusions.

Given that accounting for solar gains has been shown to be the largest source of ambiguity and uncertainty in the co-heating test, it follows that the uncertainty in the HTC will be higher in periods of higher solar irradiance, when solar gains are likely to represent a larger contribution to the total heat input. This suggests that external conditions, in particular solar irradiance levels, could impact on the accuracy of the HTC measurement by a co-heating test. In the development of a future standard test method, it may be necessary to define a set of limiting external conditions under which the test can be applied, *e.g.* as defined for blower door tests (BSI, 2001). The concept of environmental limitations upon co-heating testing has been suggested since the very early development of co-heating (Everett, 1985), and Alexander and Jenkins (2015) and Stamp (2015) have investigated these limitations more recently using simulated co-heating tests. These conditions could include maximum solar irradiance figures, but could also extend to include other parameters such as wind speed. The limiting conditions may also be intrinsically linked to the performance of the building tested, built form and length of the test (Alexander & Jenkins, 2015; Stamp, 2015). For example, measurement uncertainty will be higher for buildings with a lower HTC and so the limits on the allowable environmental conditions may be more stringent. Further research, particularly studies based upon empirical data, is required to finalize a suitable set of environmental conditions.

Solar-irradiance measurements used in regression analysis were taken in two different orientations in this project: horizontal and vertical, south-facing. The choice will affect the solar-irradiance measurements, and therefore the calculated solar aperture, solar gain and, finally, HTC. To achieve the highest possible repeatability, a common orientation for solar-irradiance measurements should be defined. Simultaneous vertically and horizontally oriented solar-irradiance measurements were not carried out in this project, so it did not generate suitable data to investigate this issue further. The Leeds Beckett method (Johnston, Miles-Shenton, Wingfield, Farmer, & Bell, 2012) recommends vertical south-facing measurements, and Stamp (2015) found that horizontal solar-irradiance measurements can lead to an overestimate of the HTC, particularly in buildings that experience high levels of direct solar gains. The results of this project show that, in this particular building and for the weather conditions occurring during the project, the orientation of the solar-irradiance measurement did not cause significant variance in the measured HTC.

No statistically significant relationship was found between reported HTC and external temperature, solar irradiance or wind speed during the testing periods, demonstrating that repeatable HTC measurements were possible during periods of different prevailing weather conditions. This is especially interesting as the testing continued beyond the traditional co-heating season and into early May, and included an unusually sunny and warm period in late March. Some teams chose to use an internal set point of greater than 25°C in order to maintain an internal–external temperature difference of at least 10°C during periods of warmer weather. Higher internal temperatures increase the chance of accelerated drying out and cracking, and care should be taken to choose a temperature that is not significantly more than 10°C above the likely local ambient temperature.

Although the 'night data only' analysis approach significantly simplifies the calculation of the HTC, it relies upon the definition of a steady-state period, independent of the influence of the preceding day's solar gain. This state seems unlikely given the high thermal time constant of many typical constructions, indeed the lack of a steady-state condition is the reason that co-heating tests are carried out over an extended period. By way of comparison, British Standards recommend that infrared thermography testing is not carried out for 12 h after a surface has been exposed to direct sunlight (BSI, 1999). Team D cited the issue of thermal storage for their selection of a sampling period that included only data

collected during the early morning (06:00–8:00 hours), but considered that even this sampling period assumed that the building was thermally very lightweight. The approach would have to be adapted for much of the typical co-heating season, where sunrise occurs before 08:00 hours.

Six of the seven teams (Table 3) accounted for the effect of varying wind speed using a multiple regression method (as described above); and the HTC is reported for zero wind speed. This may cause a problem when comparing measured and predicted HTC values because U-values are typically calculated for an assumed free wind speed of 4 m/s (BSI, 2008b). In order to allow direct comparison between predicted and measured HTC values, it may therefore be more suitable to normalize the HTC to a free wind speed of 4 m/s. The problem is that there is rarely, in the authors' experience, a significant relationship between wind speed and power consumption because wind effects are small in most cases. Thus, although wind speed does have an effect on heat transfer, it cannot be isolated from other influencing factors, a problem previously identified by Johnston et al. (2012). A linear regression method may, therefore, lead to spurious results. At present, an acceptable method to account for varying wind speeds and degrees of wind sheltering has not been developed, and it is recommended that the HTC should be reported with no adjustment for wind speed. The daily mean wind speeds throughout the test should be included in the reporting of results.

If a similar round-robin test were to be repeated, it may be useful to define the measurements and results that must be reported in order that the effects of different testing and analysis methods can be disaggregated. This would be most appropriate after a complete testing and data-analysis protocol has been documented, and would give greater insight into the repeatability of the test in practice. In the case of the project reported in this paper, such a prescriptive approach may have been inappropriate because it would have prevented the range of testing and data-analysis methods that are currently in use from being revealed. The participating teams had different levels of experience in undertaking co-heating tests, with at least one team performing their first ever test in this project. This may have contributed to some of the differences in the data-gathering and analysis methods applied; with greater experience and a standardized method it is likely that greater repeatability in the results would be achieved.

This project was performed using a single detached house, and further testing in a wider sample of houses and situations is recommended as a crucial next step towards the development of a standardized co-heating protocol. Examples of issues that this further testing should seek to address, which were not included in this study, are: the effect of attached dwellings (such as in semidetached and terraced houses and in flats or apartments); integral unheated spaces (such as garages and conservatories); overshading (of which there was little for the house studied); different glazing types (such as roof lights); suitable limiting environmental conditions; and the relationship between the accuracy of the test and the performance of the dwelling (particularly as the accuracy of the test is most typically reported as a percentage of the HTC). Each of the listed issues could add complexity to the test and cause inaccuracy in one or more of the data-analysis methods identified. Therefore, their effects must be accounted for in a comprehensive standardized protocol.

At present, the co-heating test is a useful research tool that continues to be applied on an increasingly wide basis, resulting in continual development and improvement. It would be useful, in the interim period before an agreed co-heating protocol has been finalized, to apply each data-analysis type described in this paper and to report the full set of results obtained by each one. This would allow direct comparison between tests carried out by different organizations but does not require significant further data-analysis or additional measurements (with the exception of team F's 'facade gain measurement' method). The generation of a larger database of co-heating results that can be directly compared would aid, or could be sufficient for, the development of a protocol that could become a standardized regulatory tool.

For the construction industry to adopt a carefully prescribed co-heating protocol, further development of the test would be very useful. Most notably, house builders are very reluctant to have completed properties vacant, so a test that takes two weeks between completion and occupation may be resisted. Research is underway to try to find a solution to this problem and methods that can be completed is a much shorter time frame, or which can be applied in occupied dwellings, may soon be available (Farmer, Johnston, & Miles-Shenton, 2016; Jack, 2015; Jack, Loveday, Allinson, & Porritt, 2015; Mangematin, Pandraud, & Roux, 2012; Papafragkou, Ghosh, James, Rogers, & Bahaj, 2014; Stamp, 2015). A proven and reliable test to provide benchmark values against which those of new measurement methods can be compared will be critical in the development process. The co-heating test appears best placed to provide this function.

Invariably, there is reticence about external quality control, especially if it reveals deficiencies in construction performance post-completion. What practically

could be done to rectify the problem? How much would this cost? Who pays? What are the ramifications for the other homes built to the same design? These are generic issues that would need to be resolved specifically for co-heating if it is to become a routine part of quality assurance in the house-building industry.

Conclusions

Between December 2011 and May 2012 seven independent teams conducted co-heating tests to determine the HTC of the same single detached test house located near Watford, UK. Differences in the method of testing and data analysis adopted by the teams have been described, providing a unique comparison of co-heating methods and outcomes. The robustness of the test methods is discussed and the reproducibility and variability in the measurement of HTC values is quantified.

- Of the seven testing teams, six reported final HTC values that came within ±10% of the mean (the results of the seventh team's test were excluded due to equipment failure during the test).
- Uncertainty analysis allowed an estimate of the experimental uncertainty in calculating the HTC of ±8%, based upon typical measurement and analysis accuracy and the results of a long-term co-heating testing carried out by the BRE in an adjacent, identical house. This evidence suggests that the co-heating test can be accurate to within ±8–10%.
- The largest variation in co-heating testing and analysis arises in the estimation of solar gain. When reporting results of a co-heating test, it is recommended that the method used for the estimation of solar gains should be clearly reported. Ideally, each data-analysis method described in this paper should be applied, and the results reported to allow direct comparison between tests carried out by different organizations. This will also allow the compilation of a dataset of co-heating results gathered in different situations, which will help in devising the best, standard, data-analysis protocol.
- Recommendations for a best practice co-heating testing protocol, informed by a methodological comparison carried out and reported here, have been made. These include placement of heating and air-mixing equipment in each room, the use of PID internal temperature control, and the use of back-up equipment to mitigate the effect of equipment failure.
- The limited variation observed between the teams suggests that a best-practice protocol for conducting

a co-heating test is beginning to emerge. By adopting the procedural measures identified in this paper, in addition to the guidelines set out in the Leeds Beckett University method statement (Johnston et al., 2012), repeatable data collection should be achieved.
- To ensure the widest possible applicability of a standardized co-heating protocol, it is recommended that a more comprehensive matrix of testing is set up to investigate systematically analysis procedures and measurement accuracy for a range of house types, in different locations, subject to different weather conditions.
- Finally, it is worth emphasizing that the co-heating test is crucial for quantifying the 'gap' between the actual and predicted energy performance of buildings. This gap can be of the order of 100%, so a method that is accurate to within 8–10% has a clear role to play. It is likely that the repeatability of the test could be improved given the application of a standardized best-practice protocol. This would lead towards a test suitable for compliance testing within a regulatory framework. Increased confidence in the reliability of the co-heating test could lead to its adoption and use by the industry and regulators as a quality control and compliance tool.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Richard Jack ⓘ http://orcid.org/0000-0002-5938-7858
Dennis Loveday ⓘ http://orcid.org/0000-0002-9029-0296
David Allinson ⓘ http://orcid.org/0000-0001-6161-0636
Kevin Lomas ⓘ http://orcid.org/0000-0001-5792-0762

References

Alexander, D. K., & Jenkins, H. G. (2015). The validity and reliability of co-heating tests made on highly insulated dwellings. Energy Procedia, *78*, 1732–1737. doi:10.1016/j.egypro.2015.11.282

Bauwens, G., & Roels, S. (2014). Co-heating test: A state-of-the-art. *Energy and Buildings*, *82*, 163–172. doi:10.1016/j.enbuild.2014.04.039

Bauwens, G., Standaert, P., Decluve, F., & Roels, S. (2012). Reliability of co-heating measurements. In J. Wright & M. Cook (Eds.), *Proceedings of the 2012 building simulation and optimization conference; 10-11 September 2012, Loughborough, Leicestershire, UK* (pp. 348–355). Loughborough: Loughborough University. http://doi.org/ISBN:978-1-897911-42-6

BRE. (2011). *SAP 2009 The Government's standard assessment procedure for energy rating of dwellings.* Watford: Author.

BRE. (2014). SAP 2012 The Government's standard assessment procedure for energy rating of dwellings. *Energy* (March), 174. doi.org/10.1007/s13398-014-0173-7.2

BSI. (1999). *BS EN 13187:1999 thermal performance of buildings – qualitative detection of thermal irregularities in building envelopes – infrared method.* London: British Standards Institution.

BSI. (2001). *BS EN 13829:2001 thermal performance of buildings -determination of air permeability of buildings – fan pressurization method.* London: British Standards Institution.

BSI. (2008a). *BS EN ISO 13790:2008 Energy performance of buildings – Calculation of energy use for space heating and cooling.* London: British Standards Institution.

BSI. (2008b). BS EN ISO 6946:2007. Building components and building elements – thermal resistance and thermal transmittance – Calculation method. London: British Standards Institution.

BSRIA. (2011). Building performance evaluation testing and instrumentation packages. Retrieved September 16, 2016, from https://www.bsria.co.uk/news/article/building-performance-evaluation-testing-and-instrumentation-packages/

Butler, D., & Dengel, A. (2013). Review of co-heating test methodologies. Milton Keynes: NHBC Foundation. http://doi.org/978-1-84806-351-8

CIBSE. (2015). *CIBSE guide A: Environmental design. Building* (Vol. 30). http://doi.org/10.1016/0360-1323(94)00059-2

DECC. (2011). *Digest of United Kingdom Energy Statistics (DUKES) Chapter 1: Long Term Trends.* London: Department of Energy and Climate Change.

DECC. (2012). *Energy consumption in the United Kingdom (ECUK).* London: Department of Energy and Climate Change.

Everett, R. (1985). Rapid thermal calibration of houses report ERG055. Science and engineering research council report. Milton Keynes, UK: Energy Research Group, Open University.

Farmer, D., Johnston, D., & Miles-Shenton, D. (2016). Obtaining the heat loss coefficient of a dwelling using its heating system (integrated coheating). *Energy and Buildings*, *117*, 1–10. doi:10.1016/j.enbuild.2016.02.013

Guerra-Santin, O., Tweed, C., Jenkins, H., & Jiang, S. (2013). Monitoring the performance of low energy dwellings: Two UK case studies. *Energy and Buildings*, *64*, 32–40. doi:10.1016/j.enbuild.2013.04.002

Gupta, R., Gregg, M., Passmore, S., & Stevens, G. (2015). Intent and outcomes from the retrofit for the Future programme: Key lessons. *Building Research & Information*, *43*(4), 435–451. doi:10.1080/09613218.2015.1024042

HM Government. (1998). ELECTRICITY the meters (certification) regulations 1998, *11*(1566). Retrieved from http://www.legislation.gov.uk/uksi/1998/1566/made/data.pdf

HM Government. (2006). The building regulations 2000. Conservation of fuel and power approved document part L1A (2006 edition). United Kingdom.

HM Government. (2013). The building regulations 2010. Conservation of fuel and power approved document part L1A (2013 version) United Kingdom.

IEA. (2016). IEA EBC Annex 71: Building energy performance assessment based on *in-situ* measurements. Kickoff meeting report. "Subtask 1: Gathering input data". October 10-12, 2016, Leuven, Belgium.

ISO. (1990). *ISO9060:1990 solar energy -- specification and classification of instruments for measuring hemispherical solar and direct solar radiation.*

Jack, R. (2015). *Building diagnostics: Practical measurement of the fabric thermal performance of houses.* (PhD Thesis). Loughborough University.

Jack, R., Loveday, D., Allinson, D., & Lomas, K. (2015). Quantifying the effect of window opening on the measured heat loss of a test house. In *International SEEDS Conference 2015: Sustainable Ecological Engineering Design for Society, Leeds Beckett University, Leeds, UK, 17th-18th September 2015.* Leeds.

Jack, R., Loveday, D., Allinson, D., & Porritt, S. M. (2015). Practical measurement of the thermal performance of houses. *Innovation and Research Focus* (100), 6–7. Retrieved from http://www.innovationresearchfocus.org.uk/Issues/100/IRF100_BIS.html

Johnston, D., Farmer, D., Brooke-Peat, M., & Miles-Shenton, D. (2016). Bridging the domestic building fabric performance gap. *Building Research & Information*, *44*(2), 147–159. doi:10.1080/09613218.2014.979093

Johnston, D., Miles-Shenton, D., Wingfield, J., Farmer, D., & Bell, M. (2012). Whole house heat loss test method (Coheating). In *Subtask 2: Full Scale Dynamic Testing: Ed. Gorse C. A and Erkoreka, IEA Annex 58 2nd Meeting, April 2-5 2012, Bilbao, Spain.* Retrieved from http://www.leedsmet.ac.uk/as/cebe/projects/iea_annex58/whole_house_heat_loss_test_method(coheating).pdf

Lomas, K. J., & Eppel, H. (1992). Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, *19*(1), 21–44. doi:10.1016/0378-7788(92)90033-D

Lowe, R.J., Wingfield, J., Bell, M., & Bell, J.M. (2007). Evidence for heat losses via party wall cavities in masonry

construction. *Building Services Engineering Research and Technology*, 28, 2. http://journals.sagepub.com/doi/abs/10.1177/0143624407077196

Mangematin, E., Pandraud, G., Gilles, J., & Roux, D. (2012). Patent WO 2012/028829 A1 determination of the heat loss coefficient of a premises. France.

Mangematin, E., Pandraud, G., & Roux, D. (2012). Quick measurements of energy efficiency of buildings. *Comptes Rendus Physique*, 13(4), 383–390. doi:10.1016/j.crhy.2012.04.001

Meulenaer, V. De, Veken, J. Van Der, Verbeek, G., & Hens, H. (2005). Comparison of measurements and simulations of a "passive house." In *Building simulation 20015, 9th International IBPSA Conference. Montreal, Canada August 15-18,* 2005 (pp. 769–776).

Palmer, J., Godoy-Shimizu, D., Tillson, A., & Mawditt, I. (2016). *Building performance evaluation programme: Findings from domestic projects. making reality match design*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/497758/Domestic_Building_Performance_full_report_2016.pdf

Palmer, J., Pane, G., Bell, M., & Wingfield, J. (2011). *Comparing primary and secondary terms analysis and re-normalisation (PStar) test and co-heating test results Final report – BD2702*. London: Department for Communities and Local Government.

Papafragkou, A., Ghosh, S., James, P. A. B., Rogers, A., & Bahaj, A. S. (2014). A simple, scalable and low-cost method to generate thermal diagnostics of a domestic building. *Applied Energy*, 134, 519–530. doi:10.1016/j.apenergy.2014.08.045

Roulet, C., & Foradini, F. (2002). Simple and cheap air change rate measurement using CO2 concentration decays. *International Journal of Ventilation*, 1(1), 39–44. Retrieved from http://www.ijovent.org/doi/abs/10.5555/ijov.2002.1.1.39 doi:10.1080/14733315.2002.11683620

Sherman, M. H. (1987). Estimation of infiltration from leakage and climate indicators. *Energy and Buildings*, 10(1), 81–86. doi:10.1016/0378-7788(87)90008-9

Siviour, J. B. (1981). Experimental thermal calibration of houses. In Comparative experimentation of low energy housing colloquium. University of Liege.

Stafford, A., Bell, M., & Gorse, C. (2012). Building confidence – a working paper. Retrieved from http://www.leedsmet.ac.uk/as/cebe/projects/building-confidence.pdf

Stafford, A., Johnston, D., Miles-Shenton, D., Farmer, D., Brooke-Peat, M., & Gorse, C. (2014). Adding value and meaning to coheating tests. *Structural Survey*, 32(4), 331–342. doi:10.1108/SS-01-2014-0007

Stamp, S. (2015). *Assessing uncertainty in co-heating tests: Calibrating a whole building steady state heat loss measurement method* (PhD Thesis). University College London.

Stamp, S., Lowe, R., & Altamirano-Medina, H. (2013). Using simulated co-heating tests to understand weather driven sources of uncertainty within the co-heating test method. *ECEEE 2013 Summer Study Proceedings, 3–8 June 2013, Toulon/Hyères, France*, 2049–2055. http://doi.org/ISBN978-91-980482-3-0

Stroma. (2015). Coheating testing assess the real-life performance of building fabric. Retrieved September 16, 2016, from http://www.stromatech.com/building-performance-testing/coheating-testing/

Subbarao, K. (1988). PSTAR – Primary and Secondary Terms Analysis and Renormalization: a unified approach to building energy simulations and short-term monitoring, SERI/TR-254-3175.

TSB. (2012). *Building performance evaluation, domestic buildings. guidance for project execution. Version 7*. Retrieved from https://connect.innovateuk.org/c/document_library/get_file?groupId=3270542&folderId=28283968&title=Domestic+technical+guidance+(tranche+7)

Visscher, H., Meijer, F., Majcen, D., & Itard, L. (2016). Improved governance for energy efficiency in housing. *Building Research & Information*, 44(5–6), 552–561. doi:10.1080/09613218.2016.1180808

White, J. A. (2014). *An investigation into the parameters that contribute to the gap between the designed and as-built thermal performance of British housing* (PhD Thesis). University of Nottingham.

de Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, 41, 40–49. doi:10.1016/j.autcon.2014.02.009

Wingfield, J., Johnston, D., Miles-Shenton, D., & Bell, M. (2010). Whole house heat loss test method (Coheating). Leeds: Leeds Beckett University.

Zero Carbon Hub, & NHBC Foundation. (2010). *Carbon compliance for tomorrow's new homes; Topic 4: Closing the gap between design and built performance*. London: Zero Carbon Hub.