

Upgrading the energy label for dwellings in Flanders: an example of a
behaviourally informed policy tool

Peer-reviewed author version

TARANU, Victoria; VERBEECK, Griet & NUYTS, Erik (2020) Upgrading the energy
label for dwellings in Flanders: an example of a behaviourally informed policy tool. In:
Building Research and Information, 48 (1), p. 18-33.

DOI: 10.1080/09613218.2019.1661763

Handle: <http://hdl.handle.net/1942/29780>

Upgrading the energy label for dwellings in Flanders: an example of a behaviourally-informed policy tool

Victoria Taranu^{a*}, Griet Verbeeck^a and Erik Nuyts^{ab}

^a Faculty of Architecture and Arts, Hasselt University, Kantoor E-A04, Agoralaan Gebouw E, B-3590 Diepenbeek, Belgium.

^b PXL University College, Elfde-Liniestraat 24, B-3500 Hasselt, Belgium

*victoria.taranu@uhasselt.be

Abstract

The energy performance certificate (EPC) is mandatory in the EU, aiming to enable prospective buyers and renters to compare dwellings in terms of energy performance. Flanders intends to improve its current EPC, by adding a label to the existing coloured scale and EPC score. Our research aimed to inform the Flemish Energy Agency by testing before releasing the new version of the certificate, thus being an example of behaviourally-informed policies. A literature review of experiments on information framing in similar contexts established the background to the research. Given the evidence of the importance of information framing, we tested ten versions of the new label on 224 respondents (N= 224). We verified two main hypotheses – the ‘normal distribution illusion’ and the ‘energy efficiency fallacy’. The former tested whether respondents assess too optimistically the dwellings in the red spectrum of the scale. The latter verified whether respondents focused exclusively on the energy efficiency, ignoring the dwelling size or the total energy consumption. Nudges tested to overcome these biases included social norm, anchoring and rescaling. Based on these results the label was rescaled, from the initial G–A, to F– A+. Rescaling corrects the overoptimistic assessment of the energy performance of label F.

Keywords: energy label; nudges; social norm; energy renovation; energy efficiency; laboratory experiments

Introduction

One of the information provision tools to promote energy efficiency (EE) in the residential sector is the energy performance certificate (EPC). After its introduction with the 2002/91/EC Directive (EC 2002) and reinforcement with the 2010/31/EU recast (EC 2010), it is now mandatory in the EU member states when a property is sold or rented out. Firstly, the certificate enables prospective buyers and renters to assess and compare dwellings in terms of their energy performance and to encourage them to take into account the energy performance of the property, alongside other attributes. Secondly, it aims to show the potential to improve the EE of the dwelling by presenting a set of renovation measures. In the EU countries, literature shows mixed results or limited influence of the EPC on the price of selling and renting (Mudgal, Lyons et al. 2013, Christensen, Gram-Hanssen et al. 2014, Harsman, Daghbashyan et al. 2016, Wahlstrom 2016, Marmolejo-Duarte and Chen 2019), as well as on energy renovation decisions (Christensen, Gram-Hanssen et al. 2014, Wade and Eyre 2015).

In Flanders (Belgium) the EPC scheme came into force in November 2008 for residential buildings for sale, in January 2009 for rental residential buildings (Mudgal, Lyons et al. 2013) and followed by public and non-residential buildings. The Flemish EPC had an impact, even though limited, on the market. Properties with a 1 per cent higher EPC score have a 0.075 per cent lower sales price, meaning that for two dwellings with similar characteristics, for the dwelling with an EPC score of 500 instead of 400, prospective buyers are willing to pay 4,750 euros less for a sale price of 250,000 euros (Agentschap Innoveren & Ondernemen 2016). According to the experts involved in the Individual Building Renovation Roadmap (iBROAD) project conducted by the Buildings Performance Institute Europe (BPIE), prospective buyers use the certificate to negotiate the price (BPIE 2018). On the other hand, the experts interviewed during the

iBROAD project and during our focus group, see (Taranu and Verbeeck 2017), agree that the homeowners and prospective buyers do not regard the certificate as a trustful source of information, particularly regarding renovation advice. The information is not presented in a salient way and it contains technical terms difficult to interpret. There is a range of factors that may influence the efficacy of the EPC as an information provision tool. Up to now, the supervision of the implementation of the EPC scheme at European level has mainly focused on the compliance, quality assurance and training of the certifiers (Arcipowska, Anagnostopoulos et al. 2014, Maivel, Kuusk et al. 2016). Few projects and reports analysed the importance of the content and layout (IDEAL EPBD Project, Sutherland, Audi et al. 2015), which underlined the limited research in this regard and the need for further research. In order to have the expected impact, the certificate must be elaborated in such way as to overcome individual market failures such as bounded rationality – limited capabilities to process information and maximize utility (Simon 2000, Bubb and Pildes 2014).

Previous research in behavioural sciences has shown evidence of biased interpretation of energy metrics, such as the perceived linearity of MPG (miles per gallon) (Larrick and Soll 2008, Larrick, Soll et al. 2015). Another example regards the energy label of the appliances, the ‘energy efficiency fallacy’, when consumers focus only on the energy efficiency (e.g. class A), ignoring the information about annual electricity consumption (e.g., 100 kWh/year) (Waechter, Sütterlin et al. 2015). This biased interpretation of the information on energy labels may contribute to sub-optimal decision making for consumers and ineffective information provision tools. In the recent years there is a growing interest to apply the behavioural insights regarding rational and heuristic thinking to various areas of policy making (Dolan, Hallsworth et al. 2010, Lunn 2013, Lourenço, Ciriolo et al. 2016). Yet, in the context of promoting EE

and energy renovation there are still few behaviourally-informed policies based on evidence.

The Flemish EPC contains an energy indicator, recommendations for improving the EE of the dwelling and other technical information. This present paper will focus on the EPC scheme for the residential sector, particularly on the first page containing the energy indicator. The version of the certificate that was in use in Flanders since 2008 (see Figure 1) was to be replaced in January 2019. At the time of this research the new version of the certificate was a preliminary version still under trial (see Figure 2). Based on research, including the one presented in this paper, the version launched in January 2019 has been changed, compared to the preliminary test version, see Figure 3 (Vlaamse overheid 2018). Thus, the new version of the certificate is an example of evidence-based policy that takes into account behavioural insights.

[Figure 1 near here]

[Figure 2 near here]

[Figure 3 near here]

The original EPC version had the following characteristics, see Figure 1. While most of the EU certificates contain a label, the Flemish certificate contained an EPC score and a continuous scale. The scale ranged from -50 to 750 kWh/m² per year of primary energy, with the last values specified on the scale being 0 and 700. The energy performance indicator is relative to the size of the dwelling (expressed as kWh per m² of floor area) and it is calculated for standard occupancy and a standard Belgian climate. The total primary energy was also presented on the certificate, but only on the second page of the certificate as plain text and in a less salient way than the energy label. The preliminary test version had several changes compared to the previous EPC, the main being the addition of the label in addition to the EPC score and the scale, see Figure 2.

The scale was inverted and ranged from 700 to 0 kWh/m² per year of primary energy, with the last values specified on the scale being 600 and 100.

The hypothesis of our research is that information framing plays a role in how people comprehend and interpret the preliminary new version of the Flemish energy label. A wrong interpretation of the label could mean a distorted estimation of the energy performance of a dwelling compared to the other dwellings on the market, which could indirectly impact purchase, rental or renovation decisions. By information framing we mean the following aspects: the content, the wording and the layout. We do not analyse these aspects separately because in the case of the energy indicator, the choice between EPC score and label implies changes in all these three aspects at once, as the comparison of Figure 1 and Figure 2 shows. Information framing includes framing effects: the decisions of people are affected by the wording of the same message, for example if the same statistics are presented in positive or negative terms (Kahneman and Tversky 1979), amongst other aspects of layout.

Various dual-process models state that a message can be processed in a heuristic or in a deliberative way – also called system 1/system 2 thinking (Evans and Frankish 2013). In this context biases mean cognitive errors in interpreting information, for example an optimistic estimation of the dwellings in the red spectrum due to the illusion of a uniform distribution of the existing dwellings on the scale, see label 2.1 and 2.2 of Figure 4. At the same time, biases can mean shortcuts to deliberative thinking, for example using social norm or anchoring to avoid effortful system 2 thinking (Darnton 2008). The second type of biases, if used in the right direction, become nudges that could correct the first type of cognitive biases (Baldwin 2014).

The current paper presents the results of an experimental study comprising 224 respondents (N=224) comparing ten alternative information framings of the energy

indicator. The aim was to investigate if respondents are prone to biases and if certain nudges are effective in correcting the biases. Based on a qualitative previous study (Taranu and Verbeeck 2017) the nudges tested in our study are social norm, anchoring, range effects and salience, explained in more detail in the subsection “*Aim and hypotheses of the study*”.

Importance of information framing in energy label schemes

Literature review of experimental studies

The EPC includes various messages regarding energy performance that have to be communicated and ‘translated’ to the dwellers. These messages can be illustrated with various metrics such as kWh/m² per year, U-value, CO₂ savings, pictograms, etc. The literature review below shows that the choice of the units, the wording and the layout of these indicators have an impact on how the information is processed. This section summarises a literature review on experiments that tested various information framings in energy related contexts and took into account system 1 thinking, see *Table 1*. It is not an exhaustive literature review, but a compilation of experimental studies in similar contexts - labels for appliances, water heaters and vehicles. All the studies took into account the existence of biases and heuristics in the interpretation of the label.

[Table 1 near here]

The first series of experiments presented is regarding energy labels for appliances. One of the studies has detected and validated the ‘energy efficiency fallacy’ (Waechter, Sütterlin et al. 2015). The main information on the energy label is the EE class (e.g. A+++) and the total annual electricity consumption (e.g. 100 kWh/year). The respondents were faced with one experimental condition that was one of the four combinations of low/high EE and low/high total electricity consumption. They were

asked to assess the product in terms of energy friendliness. As the EE is relative to the size of the appliance, two appliances different in size, should be compared in terms of annual electricity consumption, and not of EE class (Waechter, Sütterlin et al. 2015). Yet, the respondents based their estimations only on the EE class, largely ignoring the annual electricity consumption. This ‘energy efficiency fallacy’ might be caused by the salience of the energy class and it denotes a heuristic interpretation of the label. These findings were confirmed by experiments with eye-tracking technologies (Waechter, Sutterlin et al. 2015).

The salience effect of the class of the energy label for appliances was confirmed also by the study of Blasch et al. Due to bounded rationality, consumers use a heuristic strategy to solve the optimization problem according to a salient characteristic (price or EE class) (Blasch, Filippini et al. 2016). Another eye-tracking study has revealed bounded rationality (Waechter, Sütterlin et al. 2016). More specifically, it individuated three decision-making strategies: energy-directed lexicographic, unsystematic lexicographic and unsystematic exhaustive. These three consumer clusters had differences in identifying the optimal products in terms of energy friendliness. This example shows that it is possible to investigate also the way of reasoning, not only separate biases in isolation.

An online study focused on the EE indicator of the appliances label (Waechter, Sütterlin et al. 2016). The experimental conditions were: the original scale (A+++ to D), modified scale (A to G), modified colours and scale (A to G), shortened scale (A to C). The respondents were presented with one experimental condition. They were asked to assess the energy friendliness of two products, that had the highest EE class (e.g. A+++ on the original scale) and the second highest (e.g. A++ on the original scale). The results show that consumers might be susceptible to an anchoring effect (Furnham and

Boo 2011). The estimation of the difference in energy friendliness between the two products is affected by the number of classes of the scale, with the lowest class working as an anchor. The difference between classes was perceived to be larger for the shorter scales. Furthermore, the perceived difference is higher for A to G scale compared to A+++ to D scale due to the plus sign. The results suggest the necessity to rescale the existing EE indicator or to eliminate the plus sign.

In addition to experiments on energy labels for appliances, similar research has been undertaken regarding the US Energy Guide label (Newell and Siikamaki 2014, Newell and Siikamaki 2015). This study manipulated various information framings of the water heater labels in order to determine the individual discount rates. Alongside monetary savings, additional information was tested with alternatives such as physical energy use and carbon dioxide emissions. Yet, information on the monetary value of energy savings was the most important guiding element for choosing cost-efficient EE investments (Newell and Siikamaki 2014).

The well-known study of (Larrick and Soll 2008), shows that the choice of units, MPG (miles per gallon) instead of gallons used per 10,000 miles, can influence an erroneous understanding of the energy efficiency of vehicles. Due to the illusion of linearity of the MPG indicator, people undervalue the improvements of fuel-inefficient cars and subsequently, their replacement.

Applying behavioural insights to the EPC scheme in the EU and Flanders

The studies on labels for appliances, water heaters and vehicles of the previous subsection provide evidence of system 1 thinking in interpreting information regarding EE. These insights are complemented by the outcomes of our previous qualitative study regarding labelling schemes for residential buildings – a comparative analysis of nine European EPCs and a focus group (Taranu and Verbeeck 2017). The EPCs were

analysed taking into account behavioural insights with the aim to individuate possible information framings and nudges. The focus group with experts aimed to verify how these insights apply to the Flemish EPC and to narrow down the hypotheses for the quantitative study of the present paper. These experimental conditions will be presented in more detail in the next section.

Laboratory experiment Flemish EPC

Aim and hypotheses of the study

The aim of the study was to determine whether information framing plays a role in the comprehension and interpretation of the Flemish EPC. The first objective was to verify whether heuristic thinking interferes with deliberative thinking and whether the hypothesized cognitive biases were confirmed. These biases were 'normal distribution illusion' and 'energy efficiency fallacy'. Alongside detecting cognitive biases, the second objective of the study was to verify if nudges such as social norm, anchoring, range effects and salience are effective in correcting the above biases. All the experimental conditions are presented in Figure 4 and Table 2 summarizes the behavioural insights relative to each label.

One of the hypotheses of the experimental study is that dwellers do not interpret correctly the EPC score that is expressed in kWh/m² per year of primary energy and that is relative to the size of the dwelling. We will refer to this bias as 'energy efficiency fallacy', a bias similar to the one detected in a series of experiments regarding the energy label for the appliances (Waechter, Sutterlin et al. 2015, Waechter, Sütterlin et al. 2015). Our assumption is that respondents would focus mostly on the label and the EPC score, that represent the energy efficiency of the dwelling, ignoring the information regarding the size or the total energy.

[Table 2 near here]

Another bias that we hypothesized from the previous qualitative study regards the interpretation of the scale of the energy indicator. If we overlap the scale with the distribution of the Flemish building stock relative to each energy class, the graph is left skewed, see labels 2.1 and 2.2 in Figure 4. We suspect that in the absence of the bar graph, respondents imagine a normal distribution, therefore we will refer to this potential bias as ‘normal distribution illusion’. If confirmed, it would mean that people overestimate the percentage of dwellings with a better label than dwelling class F, having an optimistic assessment of the dwellings of the label F and a pessimistic assessment of the Flemish dwelling stock. If confirmed, this bias could possibly be corrected with specifying the median, instead of the mean, since the median value in a left skewed graph is a higher value than the mean, therefore the social norm would mean comparing with a better EPC score.

Adding references on the scale, such as the median value (label 4) or explanation of the A+ (label 6) implies anchoring manipulations, see Figure 4. These references on the scale might influence the interpretation of the energy performance of a dwelling with a certain label. When assessing a certain value, people often start from a reference point called anchor and adjust subsequently. Often the adjustment is not sufficient and the final evaluation is affected by different starting points (Kahneman, Slovic et al. 1982, Furnham and Boo 2011). Besides references on the scale, we expect that also the ranges of the scale might work as anchors and therefore, the effect of rescaling was tested with labels 5 and 6 (see Figure 4 and Table 2).

Another manipulation tested was the social norm, in the form of plain text (labels 3 and 4) or a graph (labels 2.1 and 2.2), see Figure 4 and Table 2. People are heavily influenced by others’ opinions and decisions (Brafman and Brafman 2009,

Allcott 2011, BIT 2011, Frederiks, Stennerl et al. 2015). Descriptive social norms represent the behaviour of the majority for purpose of comparison (Darnton 2008). In our study the descriptive social norm is instrumentalised as ‘70% of Flemish dwellings has a label better than F’ or ‘More than half of the Flemish dwellings have a label D or better’.

Method experimental design

The experimental research design presented in this chapter has advantage over field experiments, for example randomized controlled trials (RCTs), for being able to isolate and detect the behavioural mechanism (Lunn and Choisdealbha 2016). Moreover, there are various constraints in releasing different versions of the Flemish EPC for field experiments and detecting the impact on renovation decisions. Renovation decisions are affected by multiple factors over a long period of time and the EPC is not the only source of information. For these reasons we opted for an experimental study. Ten alternative information framings of the energy indicator were tested, grouped into seven experimental conditions due to similarities and lack of differences in results, see Figure 4 and Table 2. The study has a between subjects design, each respondent being presented with one experimental condition and the EPC of two dwellings:

- Dwelling class F - red colour, EPC score 580 kWh/m² per year
- Dwelling class C - yellow colour, EPC score 230 kWh/m² per year

The order in which the two dwellings were presented was randomised.

Respondents were students of Architecture and Interior Architecture in different years and the responses were collected in an auditorium before or after class. In total 224 complete responses were recorded, these being voluntary and not graded for the course. For the two dwellings, the first page of the certificate (with the energy indicator) and the

third page (with the recommendations) were presented on paper and the questions were filled in online on laptops and smartphones. Respondents were instructed not to look ahead in the papers provided, but they were allowed to look at the previous pages, including the pages relative to the other dwelling. Questions regarding the energy indicator were based only on the first page of the EPC, while the questions regarding the recommendations were based on both pages of the certificate. In this paper we present the results regarding the first page of the certificate, the one including the energy indicator, see Figure 2.

The dependent variables can be classified in two main categories:

- Absolute terms (performance of the dwelling on its own)
 - Perceived EE
 - Perceived energy consumption
- Relative terms (performance of the dwelling compared to the rest of the Flemish building stock)
 - Percentage of dwellings with a better label

The first category of questions in the online questionnaire refers to a subjective assessment of the performance of the dwelling by each individual. Respondents had to rate the dwelling in terms of energy efficiency on a Likert scale from 1 to 5, 1 being ‘very inefficient’ and 5 ‘very efficient’. Then, they were asked “How much energy do you think an average Flemish family would consume if they would live in this house?” the Likert scale being from ‘very few energy’ to a ‘lot of energy’.

Because the energy consumption is calculated for an average Flemish family, it refers to the energy performance of the dwelling as such, excluding the consumption pattern of the users. Thus, the answers regarding the EE and energy consumption should

be the same but inverted. The duplication had the purpose to see which concept is understood better and whether there is consistency in the answers.

For the second category of dependent variables, respondents had to assess the dwelling relative to the rest of the building stock – they had to estimate the percentage of Flemish dwellings with a better energy class. If the previous Likert scale is subjective and might be different for each individual, this assessment can be compared with objective figures – 70 per cent for class F and 20 per cent for class C, according to EPC data from 2014 (Verbeeck and Ceulemans 2016). The aim of this question was to verify the hypothesis of a ‘normal distribution illusion’ and whether respondents were assessing too optimistically the red spectrum classes.

Finally, for detecting the ‘energy efficiency fallacy’, respondents had to choose which dwelling would have higher energy bills if the same family would live in:

- a dwelling with size of 100 m², class F and EPC score of 580 kWh/m² per year
- a dwelling with size of 300 m², class C and EPC score of 230 kWh/m² per year

The energy indicators of the two dwellings were presented, alongside the information of the dwelling size.

[Figure 3 near here]

Method data analysis

The differences between labels 1.1, 1.2 and 1.3 are very small. For every test we did, no significant difference between them could be found. Therefore, they are presented as the same experimental condition, label 1. Comparably labels 2.1 and 2.1 are presented as the same experimental condition, label 2. This reduces the ten test conditions (Figure 4) into 7 energy labels (Table 2).

The statistical analysis was computed using RStudio version (2016). For assessing the impact of the experimental conditions on the perceived energy efficiency and energy consumption, ANOVA Kruskal-Wallis tests and ordinal logistic regressions were performed. The ANOVA Kruskal-Wallis test is suitable for ordinal variables such as the Likert scale values. The ordinal logistic regression model has a parallel regression assumption implying that for all levels of the dependent variable the odds ratios of all the pairs of responses of the independent variable are assumed to be the same. The advantage of using an ordinal model is that it does not assume that differences between the levels of the response variable are equal, for example, the difference between “very inefficient” and “inefficient” might be different from that between “inefficient” and “average”.

For the comparison with the rest of the Flemish building stock, firstly we performed a one sample t-test to compare the estimation with the actual value for dwelling F (70 per cent) and dwelling C (20 per cent). One sample t-tests treat the discrete variable (the estimation of the percentage of dwellings performing better was measured with a definition of 10 per cent) as a continuous variable. If n is large enough ($n \geq 30$) then by the Law of Large Numbers, no assumption concerning the distribution is required. Secondly, we compare the estimation between different experimental conditions with an ANOVA Kruskal-Wallis test by ranks.

Main findings

Perceived energy efficiency

The first objective was to verify whether various ways of presenting the information (the seven energy labels) had an impact on how respondents interpret and assess the EE of the dwelling. The box plots from Figure 5 and Figure 6 illustrate the lower quartile,

the median and the upper quartile for the estimated EE of dwelling C and F relative to the experimental conditions. When performing an ANOVA Kruskal-Wallis test by ranks, for a dwelling class C, the estimated EE differs significantly depending on the way the information was presented, see Table 3. For the dwelling in the red spectrum labelled F, no significant differences are found for EE since respondents tend to agree that it is a very inefficient dwelling (see Figure 6).

[Table 3 near here]

Subsequently we compared the estimated EE of different labels with control condition E1 using the ordinal logistic regression (Proportional-odds cumulative-logit model), see Table 4. For the yellow spectrum dwelling class C, respondents assess the EE less optimistically seeing labels 5 and 6 compared with control condition label 1. All other comparisons gave no significant differences. Labels 5 and 6 are the rescaled experimental conditions: instead of the scale from class G to A (700 to 0) of the label 1 (control condition), label 5 has a scale ranging from class F to A (600 to 0) and label 6 a scale from class F to A+ (600 to -100), see Figure 4. It could be explained by the anchor effect, where the ranges of the scale work as anchors that affect the interpretation of the EPC label and score. Again, for the dwelling class F, no significant differences could be found. The results show that the order in which the dwellings were presented to the respondent influences the estimated EE. The ordinal logistic model controls for order effects by including the order variable besides the experimental conditions.

[Table 4 near here]

[Figure 4 near here]

[Figure 5 near here]

Perceived energy consumption

The results of the ANOVA Kruskal-Wallis tests show that for a dwelling class C the

estimated energy consumption differs significantly depending on the way the information is presented, see Table 3. For the dwelling in the red spectrum labelled F, no significant differences are found, respondents tend to agree that it consumes a lot of energy (see Figure 8).

The results of the ordinal logistic regression of the energy consumption confirm the results of the estimation of the EE, with a less optimistic assessment of dwelling C when presented with labels 5 and 6 compared to control condition 1. Additionally, when presented with label 7, they assess only the energy consumption higher than label 1 (see Table 4. Labels 5 and 6 are the rescaled experimental conditions, while label 7 contains the information regarding the total energy as plain text, see Figure 4.

Another novelty of the results of the energy consumption in comparison with the EE and the previous ANOVA tests lies in the perceived energy consumption of the dwelling in the red spectrum, labelled F, that is significant for label 6 (see Table 4. Also in this case, rescaling as in the experimental condition 6 (from class G to A+) contributes to a less optimistic assessment of the energy consumption of the dwelling.

For both dwellings and for both the EE estimation and the energy consumption, the order in which the dwellings were presented played a significant role. If presented second, the energy consumption of dwelling F is perceived to be higher due to the anchoring effect by contrast.

[Figure 6 near here]

[Figure 7 near here]

Comparison with the rest of the Flemish building stock

In the previous subsection we discussed the findings regarding assessing the dwellings in absolute terms, compared to the most and least EE dwelling imagined. Yet, these scales are subjective, therefore another type of comparison is needed. The respondents

were asked to estimate the percentage of Flemish dwellings with a better label than the dwelling presented. This estimation can be compared with an objective value of the existing building stock, that is 70 per cent for the dwelling labelled F and 20 per cent for the dwelling with label C, according to data from 2014 (Verbeeck and Ceulemans 2016).

Firstly, we verify if respondents have estimated correctly this value. Table 5 summarizes the results of one sample t-tests, comparing the estimation with the actual value for dwelling F (70 per cent) and dwelling C (20 per cent). For the dwelling F, the mean difference for Labels 1 (control condition), 4, 5, and 7 is significant, therefore the estimation is wrong. With labels 2, 3, and 6 the estimation of the percentage of dwellings performing better was not significantly wrong. Two of the experimental conditions present this information explicitly, either in the form of a bar graph (label 2) or as plain text (label 3). The interesting finding is that by rescaling (label 6), respondents could estimate the percentage of dwellings performing better than the dwelling in the red spectrum (label F) more accurately than with the control condition. The rescaling of label 6 means having the scale from class F to A+ instead of class G to A of label 1. With the control condition, respondents underestimated the percentage of dwellings with a better label than F. This implies a too optimistic assessment of the energy performance of the dwelling labelled F or a too pessimistic assessment of the rest of the Flemish building stock or both. This did not occur with label 6.

As for the dwelling class C, respondents had difficulties to estimate the correct value, independently of the experimental condition, see Table 5. For all labels, respondents overestimated the percentage of the dwellings performing better than class C. This implies both a too pessimistic assessment of dwellings labelled C and a too

optimistic assessment of the Flemish buildings in the yellow-green spectrum of the scale.

Secondly, we compare the estimation between different experimental conditions with the ANOVA Kruskal-Wallis test by ranks. Only for the estimation of the percentage of dwellings better than class F the differences between experimental conditions are significant.

[Table 5 near here]

The initial hypothesis was that people imagine a normal distribution on the scale of the percentages relative to the classes, instead of the left skewed distribution, see label 2.1 and 2.2 of Figure 4. This would imply that for the dwellings class F and C respondents would overestimate the percentage of dwellings with a better label. However, the results show this overestimation only for the dwelling labelled C, with the overall mean of the estimation at 35 per cent compared to 20 per cent in reality. For the dwelling class F, the estimated value was as an average lower than the actual figure, 60.4 per cent compared to 70 per cent in reality. That could imply that individuals imagine a uniform distribution on the scale instead of a normal distribution as hypothesized.

[Figure 8 near here]

[Figure 9 near here]

[Table 6 near here]

Energy efficiency fallacy

At the end of the survey, respondents had to choose which house would mean higher energy bills, given that the same family would live in both – a dwelling with class F (EPC score 580 kWh/m² per year) and size 100 m² or a dwelling with class C (EPC score 230 kWh/m² per year) and size 300 m². The purpose of this question was to verify

if respondents were aware that the EPC score is relative to the size and that the dwelling C, even though with a better EE, would require higher total energy use due to the size (69.000 kWh per year versus 58.000 kWh per year). Even though the respondents were students of the Faculty of Architecture, 72 per cent of the answers was wrong. This result confirms the hypothesis of the 'energy efficiency fallacy'. The label 7 that includes the information about the total energy consumption did not help to overcome this bias. It is important to notice that the label 7 was not presented on the same page with the last question, but it was only present for the previous questions, even though the pages were available for looking back.

Discussion - Policy implications

In many policy areas including energy efficiency, the laboratory experimental research is underused compared to field experiments, such as RCTs (Lunn and Choidealbha 2016). Nevertheless, there are many contexts where RCTs are not applicable. Renovation decisions for instance are difficult to track over time, since many recommended measures do not need a building permit. Moreover, the cause of such difficult decisions cannot be separated from complex reality and the EPC is only one of the many sources of information in play. The same is valid for home purchase decisions. Therefore, policy instruments that are aimed to influence renovation and home purchase decisions should be tested differently. Laboratory experiments and similar techniques can inform policy makers with applicable insights before the policy instrument is launched and scaled up. The advantage of laboratory experiments over field experiments is the possibility to isolate, detect and validate the behavioural mechanism behind the success of a control condition. The current laboratory study has the limitation of not differentiating the willingness from the ability to process the information since no incentive was provided to respondents for correct answers. Yet,

this lack of incentive is more similar to real-life setting when homeowners are presented with the certificate.

Another limitation of the laboratory study is the sample of Architecture students, who potentially have better knowledge regarding the energy performance of the buildings. Yet, 72 per cent was affected by the ‘energy efficiency fallacy’ and answered the question wrongly. On average, general public would perform even worse at interpreting graphs. The bar graph of label 2 (see Figure 4) was interpreted correctly by the students and helped estimating the percentage of dwellings. Hence, the positive effect of label 2 on estimating the percentage of dwellings performing better than label F (table 5) might be lost with a general public.

Results indicate a ‘uniform distribution illusion’ instead of a ‘normal distribution illusion’ initially hypothesised. Respondents are imagining the percentage of dwellings relative to each class is a uniform distribution, while in reality it is left skewed, see label 2 of Figure 4. With the control version, individuals assess too optimistically the dwelling in the red spectrum (labelled F) and underestimate the percentage of Flemish dwellings with a better label. This wrong perception could influence the purchase and renovation decisions in a negative way. Yet, this wrong perception can be corrected either by explicitly stating the percentage of dwellings performing better either as plain text (label 3) or as a bar graph (label 2), or by rescaling (label 6). The ranges of the scale work as anchors and rescaling contributes to a less optimistic assessment of the energy consumption of the label F and a correct estimation of the percentage of dwellings with a better label. The efficacy of rescaling and the anchor effect is confirmed by previous studies regarding energy labels for appliances (Waechter, Sütterlin et al. 2016). The use of the median value (label 4 in Figure 4) proved not to be effective. In the yellow spectrum, on the other hand, individuals assess

too pessimistically the dwelling with label C and overestimate the percentage of dwellings doing better. None of the experimental conditions corrected this bias, but its effect on the purchase and renovation decisions theoretically is not as negative as for the red spectrum.

The second hypothesis of an ‘energy efficiency fallacy’ was confirmed, as the interpretation of the control condition label 1 was prone to biased interpretation. The vast majority of the respondents (72 per cent) is not aware that the EPC score is relative to the size. A similar bias was detected in previous studies regarding energy labels for appliances (Waechter, Sütterlin et al. 2015). In both cases individuals focus only on the energy class, ignoring other information, such as total energy use or size. Although the total energy consumption is specified on the label for appliances, in the case of the Flemish EPC neither the value of the total energy use nor the size of the dwelling is presented on the first page of the certificate next to the energy indicator. At the same time, stating in text the total energy consumption did not prove to be effective and other ways of communicating this message should be tested.

In recent years there is a growing application of behavioural insights to policy making, with many governmental agencies collaborating with researchers or even setting up their own behavioural units (Halpern and Sanders 2016). Nevertheless, behavioural studies show contradictory results, depending on the context, or some hypothetical nudges can even backfire. The impact of certain policy instruments such as the EPC are prolonged over time, as the previous version of the Flemish EPC was in use for almost 10 years and the certificates released are valid for another 10 years. Therefore its comprehension and interpretation are worth testing before the release.

The quantitative experimental studies have to be based on previous qualitative analysis, such as focus groups, that allows exploring the implications of theoretical

behavioural insights to this particular context. The present research, is an example of how researchers can collaborate with policy makers for elaborating evidence-based policies. For most of the parameters measured, the alternative design, label 6, showed better results than the control condition label 1. Based on these results, the energy label of the new Flemish EPC was rescaled to classes F to A+ (600 to -100 kWh/m² per year) instead of the initial intended G to A (700 to 0 kWh/m² per year), see Figure 3.

Conclusions

Previous studies regarding the EPC scheme focused on technical aspects and its impact on the market. The efficacy of the EPC policy tool implies that prospective buyers and renters are able to process the technical information in a rational way. On the other hand, behavioural sciences evidence bounded rationality, a biased interpretation of energy metrics and the importance of information framing. Yet, these findings were not taken into account for the elaboration and evaluation of most of the European EPC schemes. The present study verified in a quantitative way the impact of information framing of the new version of the Flemish EPC on comprehension and interpretation of the energy indicator.

The ‘energy efficiency’ fallacy was confirmed for the Flemish EPC, even though presenting the total energy consumption as plain text next to the EE label was not effective to correct the bias. Other ways of presenting the information that the EPC score is relative to the size of the dwelling should be investigated.

Findings indicate that individuals are also prone to a ‘uniform distribution illusion’ that can be corrected with stating the percentage of dwellings with a better label (as a bar graph or plain text) or by rescaling the EPC indicator scale from the initial G-A to F- A+. Ranges of the scale have an ‘anchor’ effect that influences and

corrects the interpretation of the EE of a dwelling compared to the rest of the building stock.

Often the application of behavioural insights is associated exclusively to nudging. However even if nudging is not the purpose, various policies including mandates such as the EPC should be tested for behavioural effects before upscaling. No information is neutral and this study confirms it, since without a prior testing, the scale of the energy indicator might have contributed to a too optimistic interpretation of the energy performance of the dwellings that should be the target of renovation policies. With a correct interpretation of the energy indicator the prospective buyer and renters can make informed purchase and renovation decisions.

Declaration of interest statement: 'No conflict of interest'.

References

- Agentschap Innoveren & Ondernemen (2016). RENBEN Sociaal-maatschappelijk Rapport [RENBEN Sociological Report]: 44.
- Allcott, H. (2011). "Social norms and energy conservation." Journal of Public Economics **95**(9): 1082-1095.
- Arcipowska, A., F. Anagnostopoulos, F. Mariottini and S. Kunkel (2014). Energy Performance Certificates Across the EU. A mapping of national approaches., BPIE: 60.
- Baldwin, R. (2014). "From Regulation to Behaviour Change: Giving Nudge the Third Degree." Modern Law Review **77**(6): 831-857.
- BIT (2011). Behaviour Change and Energy Use, the Cabinet Office Behavioural Insights Team: 35.
- Blasch, J., M. Filippini and N. Kumar (2016). "Boundedly rational consumers, energy and investment literacy, and the display of information on household appliances." ETH-Zürich Economics working paper series 16/249.
- BPIE. (2018). "iBROAD Project." Retrieved April 2018, from <http://bpie.eu/publication/current-use-of-energy-performance-certificates-in-8-eu-countries-and-potential-links-to-ibroad/>.
- Brafman, O. and R. Brafman (2009). Sway : the irresistible pull of irrational behaviour. London, Virgin.
- Bubb, R. and R. H. Pildes (2014). "How behavioural economics trims its sails and why." Harvard Law Review **127**(1593): 1594-1678.

Christensen, T. H., K. Gram-Hanssen, M. de Best-Waldhober and A. Adjei (2014). "Energy retrofits of Danish homes: is the Energy Performance Certificate useful?" Building Research & Information **42**(4): 489–500.

Darnton, A. (2008). Reference report: an overview of behaviour change models and their uses, GSR Government Social Research.

Dolan, P., M. Hallsworth, D. Halpern, D. King and I. Vlaev (2010). MINDSPACE Influencing behaviour through public policy, Institute for Government: 96.

EC (2002). Directive 2002/91/EC on the energy performance of the buildings The European Parliament and the Council of the European Union. Official Journal of the European Communities.

EC (2010). Directive 2010/31/EU on the energy performance of buildings (recast). The European Parliament and the Council of the European Union. Official Journal of the European Union.

Evans, J. S. B. T. and K. Frankish (2013). In two minds: dual processes and beyond. Oxford, Oxford University Press.

Frederiks, E. R., K. Stennerl and E. V. Hobman (2015). "Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour." Renewable & Sustainable Energy Reviews **41**: 1385-1394.

Furnham, A. and H. C. Boo (2011). "A literature review of the anchoring effect." Journal of Socio-Economics **40**(1): 35-42.

Halpern, D. and M. Sanders (2016). "Nudging by government: Progress, impact & lessons learned." Behavioural Science & Policy **2**(2): 53-65.

Harsman, B., Z. Daghbashyan and P. Chaudhary (2016). "On the quality and impact of residential energy performance certificates." Energy and Buildings **133**: 711-723.

IDEAL EPBD Project. "IDEAL EPBD Project Improving Dwellings by Enhancing Actions on Labelling for the EPBD." May 2016, from <https://ec.europa.eu/energy/intelligent/projects/en/projects/ideal-epbd>.

Kahneman, D., P. Slovic and A. Tversky (1982). Judgment under uncertainty: Heuristics and biases, Cambridge University Press.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." Econometrica **47**(2): 263-291.

Larrick, R. P. and J. B. Soll (2008). "Economics - The MPG illusion." SCIENCE **320**(5883): 1593-1594.

Larrick, R. P., J. B. Soll and R. L. Keeney (2015). "Designing better energy metrics for consumers." Behavioral Science & Policy **1**(1): 63-75.

Lourenço, J. S., E. Ciriolo, S. R. Almeida and X. Troussard (2016). Behavioural Insights Applied to Policy. European Report 2016, European Commission. Joint Research Centre JRC: 54.

Lunn, P. (2013). Behavioural Economics and Regulatory Policy, Public Governance And Territorial Development Directorate. Regulatory Policy Committee.

Lunn, P. D. and Á. N. Choisedalbha (2016). "The case for laboratory experiments in behavioural public policy." Behavioural Public Policy.

Maivel, M., K. Kuusk, R. Simson and J. Kurn (2016). Overview of existing surveys on energy performance related quality and compliance, QUALICHeCK Project: 104.

Marmolejo-Duarte, C. and A. Chen (2019). "La incidencia de las etiquetas energéticas EPC en el mercado plurifamiliar español: un análisis para Barcelona, Valencia y Alicante [The impact of EPC rankings on the Spanish residential market: an analysis for Barcelona, Valence and Alicante]." CIUDAD Y TERRITORIO. ESTUDIOS TERRITORIALES.(199): 101-118.

Mudgal, S., L. Lyons, F. Cohen, R. Lyons and D. Fedrigo-Fazio (2013). Energy performance certificates in buildings and their impact on transaction prices and rents in selected EU countries, Final report prepared for European Commission (DG Energy): 158.

Newell, R. G. and J. V. Siikamaki (2014). "Nudging Energy Efficiency Behavior: The Role of Information Labels." Journal of the Association of Environmental and Resource Economists **1**(4).

Newell, R. G. and J. V. Siikamaki (2015). "Individual Time Preferences and Energy Efficiency." AMERICAN ECONOMIC REVIEW **105**(5): 196-200.

Simon, H. A. (2000). "Bounded rationality in social science: Today and tomorrow." Mind & Society **1**(1): 25-39.

Sutherland, G., P. G. Audi and A. Lacourt (2015). 2016 Implementing the Energy Performance of Buildings Directive (EPBD), CA EPBD III.

Taranu, V. and G. Verbeeck (2017). "A closer look into the European Energy Performance Certificates under the lenses of behavioural insights—a comparative analysis." Energy Efficiency.

Verbeeck, G. and W. Ceulemans (2016). Samenvattend rapport analyse van de EPC databank. Resultaten tot en met 2014. [Summary report of analysis of the EPC database. Results up to and including 2014]. Leuven, Steunpunt Wonen: 55.

Vlaamse overheid. (2018). "Energiesparen. Het nieuwe EPC vanaf 2019. [Energy savings. The new EPC starting from 2019.]." 2018, from <http://www.energiesparen.be/het-nieuwe-epc-vanaf-2019>.

Wade, J. and N. Eyre (2015). Energy Efficiency Evaluation: The evidence for real energy savings from energy efficiency programmes in the household sector, UKERC UK Energy research Centre: 68.

Waechter, S., B. Sütterlin, J. Borghoff and M. Siegrist (2016). "Letters, signs, and colors: How the display of energy-efficiency information influences consumer assessments of products." Energy Research & Social Science **15**: 86-95.

Waechter, S., B. Sutterlin and M. Siegrist (2015). "Desired and Undesired Effects of Energy Labels-An Eye-Tracking Study." PLOS ONE **10**(7): e0134132.

Waechter, S., B. Sütterlin and M. Siegrist (2015). "The misleading effect of energy efficiency information on perceived energy friendliness of electric goods." Journal of Cleaner Production **93**: 193-202.

Waechter, S., B. Sütterlin and M. Siegrist (2016). "Decision-Making Strategies for the Choice of Energy-friendly Products." Journal of Consumer Policy: 1-23.

Wahlstrom, M. H. (2016). "Doing good but not that well? A dilemma for energy conserving homeowners." Energy Economics **60**: 197-205.

Table 1 Overview of studies of information framing effects in energy related contexts

Reference	Context	Method	Sample	Behavioural insights
(Waechter, Sütterlin et al. 2015)	Energy labels appliances	Online experiment	169	‘Energy efficiency fallacy’.
(Waechter, Sutterlin et al. 2015)	Energy labels appliances	Eye tracking experiment	117	Bounded rationality. ‘Energy efficiency fallacy’.
(Waechter, Sütterlin et al. 2016)	Energy labels appliances	Online experiment	Study 1 - 217; Study 2 - 330	Anchor effect – ranges of the scale
(Waechter, Sütterlin et al. 2016)	Energy labels appliances	Eye tracking experiment	59	Bounded rationality
(Blasch, Filippini et al. 2016)	Energy labels appliances	Online randomized controlled choice experiments	583; 877; 1,375	Bounded rationality Salience
(Larrick and Soll 2008)	Fuel economy labels	Study 1 and 2 - experimental studies Study 3- online survey	Study 1- 77; Study 2 - 74; Study 3 - 171	Linear perception of the metric MPG (miles per gallon)
(Newell and Siikamaki 2014, Newell and Siikamaki 2015)	US Energy Guide water heaters label.	Online choice experiments and randomized information treatments	1,217	Discounting the future.

Table 2 Description of the experimental conditions of the study.

<i>Experimental condition</i>	<i>Description</i>	<i>Behavioural insights</i>
Energy label 1	Control condition. Scale from 700 to 0. References – 3 steps of renovation, EPC score average of same type of dwellings, Flemish policy goal for 2050.	Anchoring, Social norm, Information overload.
Energy label 2	Bar graph – percentage of Flemish dwellings corresponding to all the labels.	Social norm. Salience.
Energy label 3	Text – percentage of Flemish dwellings with a better label	Anchoring. Social norm.
Energy label 4	Text – Reference on the scale of the median of the Flemish building stock	Anchoring. Social norm.
Energy label 5	Re-scale from G to A (700 to 0 kWh/m ² per year) to F to A (600 to 0 kWh/m ² per year)	Anchoring - range effects.
Energy label 6	Re-scale from G to A (700 to 0 kWh/m ² per year) to F to A+ (600 to - 100 kWh/m ² per year). Text explanation of A+.	Anchoring. Range effects.
Energy label 7	Text – Total energy consumption.	Correcting the ‘energy efficiency fallacy’.

Table 3 Perceived energy efficiency and energy consumption.

	Dwelling C			Dwelling F		
	chi-squared	df	p-value	chi-squared	df	p-value
Estimated energy efficiency	17.3	6	0.008**	8.0876	6	0.2318
Estimated energy consumption	21.1	6	0.002**	7.6842	6	0.2622
Significance	***p-value < 0.001; **p-value < 0.01; *p-value < 0.05					

Table 4 Perceived energy efficiency and perceived energy consumption. Results of Proportional-odds cumulative-logit model.

	Perceived EE dwelling C	Perceived EE dwelling F	Perceived energy consumption dwelling C	Perceived energy consumption dwelling F
	Estimate	Estimate	Estimate	Estimate
(Intercept):1	-6.413***	0.684*	-3.379***	-4.836***
(Intercept):2	-3.163***	2.228***	0.3402	-2.733***
(Intercept):3	-0.575*	2.695***	2.952***	-1.422***
(Intercept):4	3.421***	3.529***	7.243***	0.623*
QE2	0.482	0.0103	-0.523	-0.169
QE3	0.075	0.2165	-0.224	-0.551
QE4	0.0095	-0.272.	-0.328	0.1317
QE5	1.055*	1.719.	-1.468**	-0.750
QE6	1.555***	1.338	-1.516**	-1.1687*
QE7	0.3475	0.47	-1.681***	-0.310
ORDER CF	0.772**	0.74*	-1.476***	-0.920***
Significance	***p-value < 0.001; **p-value < 0.01; *p-value < 0.05			

Table 5 Estimated percentage of dwellings with a better label. Comparison with the data of the building stock.

	Dwelling labelled F	Dwelling labelled C
	Mean¹	Mean²
Label 1	-15.86***	13.43***
Label 2	-3.96	14.17***
Label 3	-6	8**
Label 4	-10.95*	18.09***
Label 5	-12.22*	18.89***
Label 6	-4.58	21.25***
Label 7	-7.39*	15.65***
¹ t.test H0 mean diff (X-70) is equal to 0 ² t.test H0 mean diff (X-20) is equal to 0 Significance ***p-value < 0.001; **p-value < 0.01; *p-value < 0.05		

Table 6 Estimated percentage of dwellings with a better label. Comparison with the control condition, label 1.

	ANOVA Kruskal-Wallis p-value
Dwelling labelled F	0.0298*
Dwelling labelled C	0.0729
Significance ***p-value < 0.001; **p-value < 0.01 ; *p-value < 0.05	

Figure 1 First page of the existing version of the Flemish EPC in use until end of 2018. The energy performance indicator consists of an EPC score and a continuous scale.

Figure 2 New version of the Flemish EPC. The energy performance indicator consists of an EPC score, a label placed on a continuous scale.

Figure 3 Upper - preliminary energy label (control condition for the study). Lower - final energy label of the Flemish EPC after the study.

Figure 4 Experimental conditions.

Figure 5 Perceived energy efficiency - Dwelling labelled C

Figure 6 Perceived energy efficiency - Dwelling labelled F.

Figure 7 Perceived energy consumption - Dwelling labelled C.

Figure 8 Perceived energy consumption - Dwelling labelled F.

Figure 9 Estimated percentage of dwellings with a better label than the dwelling labelled C.

Figure 10 Estimated percentage of dwellings with a better label than the dwelling labelled F.