A multi-country comparative evaluation of labelling research

Christine Egan
Collaborative Labeling and Appliance Standards Program (CLASP)
1200 18th St., Suite 900
Washington DC, 20009 USA
cegan@clasponline.org

Paul Waide
International Energy Agency (IEA)
9 rue de la Federation
75739 Paris Cedex 15 France
paul.waide@iea.org

Keywords

Labelling, energy label design, consumer and market research, optimised market transformation

Abstract

Some 73 countries comprising a combined population of 4.9 billion have or are implementing energy efficiency information labels for appliances. Some labels have been designed through research and some have not. This paper presents a comparative evaluation of the market research completed to design new, and evaluate existing, energy information labelling programs in: Australia, China, the European Union (EU), Ghana, India, Mexico, South Africa, Tunisia and the USA. Taking into account the methodological approaches adopted the paper: a) catalogues individual findings; and b) synthesises common lessons learned across the varied projects. Results that appear to be true regardless of cultural or economic context are emphasised to arrive at "common truths" of labelling design and evaluation research.

The findings provide documentation and evidence of the importance of conducting consumer market research when designing a new energy label or modifying an existing one. Some spectacular actual or potential failures are highlighted when this has not been the case. The paper serves as an inventory of work in the field of information labelling design research as well as a best practice guide for policymakers interested in undertaking evaluations of existing labelling schemes.

Introduction

A wide range of energy labels are presently in use globally. Many have been designed without research to test effectiveness. An energy label is a vehicle for conveying information to consumers about the energy performance of appliances. To be effective as a market transformation instrument, it must provide that information in a way that positively influences consumer purchase decisions. There is a growing appreciation that the most influential label design is not likely to be found via an ad-hoc, committee-driven process, but rather should be determined by research focused on the primary end-user—the consumer.

This paper reports on work to date regarding label research. For organizational purposes, the paper splits the available research into four categories: (1) research to design a label from first principles: (2) research to inform an a priori label selection: (3) hybrids of the first two: and (4) research to amend an existing label.

Research that helped design the label from first principles

Research to design a label design from first principles has been conducted in the EU, India and China. The same type of research has also been done in the USA but has not yet been applied to the energy label utilised (Egan 2000). (See below, section "Research to consider an amendment to an existing label").

EUROPEAN UNION

The current European energy labelling programme was launched in 1992 with the publication of a framework Directive on energy labelling mandating the European Commission (EC) to develop compulsory energy labelling regulations for specific household appliances. Refrigerators and freezers were the first products to be labelled following the issue of an implementing Directive in January 1994. This mandatory energy labelling scheme replaced an earlier European Community scheme, which gave Member States discretion to require labelling of some household appliances sold within their jurisdiction providing a common EU label format was used (Figure 1c). This earlier label was of an information-only type, which presented technical details including energy consumption under standard test conditions, but not information about the appliance's relative energy-performance or efficiency compared to similar models. In practice, this label was only briefly applied in one Member State (Denmark) for one appliance (household ovens) and hence had a negligible impact. The simultaneous creation of the European Single Market and the desire by some Member States to develop a more meaningful labelling scheme led to the creation of a single, harmonised and mandatory label across the entire EU.

The World's first energy label design research

Early in the process, the Dutch energy and environment agency, Novem, in conjunction with the EC realised that a new EU label should be based on the findings of consumer research. The University of Leiden (UoL) and a graphic designer conducted a series of consumer research experiments to test potential labelling concepts. The research was conducted in three tranches from 1991 to 1993 (University of Leiden 1991, 1992, 1993) and drew upon earlier research which had established that appliance energy consumption should be presented relative to that of other comparable appliances. The first experiment involved testing five label designs (Figure 1a-e) that comprised: a) the existing EC label; b) a design based on the US EnergyGuide label; c) a design loosely based on the Australian categorical label using stars; d) the 'Leiden Horizontal' label; and e) the 'Leiden Vertical' label. The last two labels drew upon earlier research by UoL. All labels except the EC label included information on both the product's absolute energy consumption and its consumption relative to similar products. Labels exhibiting energy operating cost were considered, but discarded because so many differing tariffs were in place across Europe.

Figure 1. *The first five label designs including the old EC label.*

The designs were tested in a twin experiment. In the first round, 75 people were each shown the five labels for one minute individually via slides. Questions were posed through a standard tape recording. Labels were shown in a strict rotation so each label was equally seen first, second, third, etc. Respondents evaluated the labels for seven parameters on an 11 point bi-polar scale. Responses were then grouped to give scores for four major parameters: comprehension, salience, information and appeal. The difference in average scores for the labels under each of these was not statistically significant except: a) the old EC label was significantly less salient than the others and thus might be ignored and b) the Leiden Star Label was thought to be less informative than the other labels.

In addition to these results, the ability of the respondents to remember the energy consumption value (the so-called "recall" level) was tested quantitatively. The Leiden Vertical label was found to have a significantly higher level of recall compared to the modified US Energy Guide label, which scored poorly. The second part of the experiment examined how fast and accurately information was retrieved from the labels. The speed vs. accuracy trade off was deemed important for label effectiveness as earlier research found that when consumers choose between models they use a two-stage process. Initially, people tend to compare all alternatives on one or a few attributes (e.g. price, dimensions, etc.), afterwards they compare a few products on all available attributes. It is therefore important that the information given in an energy label be "top-of-mind" in the both stages. A sample of 60 people were asked to identify the most efficient, next most efficient, next to least efficient and least efficient appliance from a slide showing four cases of the same label i.e. the same design for four appliances with different energy levels were shown without any time limit. The results showed significant response time differences. On average, it took 4.5 seconds for the EC label, 3.1 seconds for the US and Leiden Horizontal labels, 3 for the Leiden Star label and 2.8 for the Leiden Vertical label. The accuracy scores were very similar between all labels (about 85%). Based on the above, the Old EC label, US-label and Leiden Star label were excluded from subsequent research.

The second exercise involved a five expert review as follows: a) an expert graphic design (with a history in designing banknotes, vouchers and street signs), b) an expert in marketing communication and consumer behaviour, c) an expert in human cognitive processes and memory, d) an expert in the labelling of hazardous substances, and e) an expert in graphic design and information ergonomics. These experts were invited to comment on specific features of the Leiden-Horizontal and Leiden-Vertical label, both of which used a thermometer concept to indicate relative energy performance. This process lead to the following main recommendations:

- Use an arrow to make a clear connection between the number of kWh and the thermometer reading
- Standardise font sizes highlighting important and minimising less important information

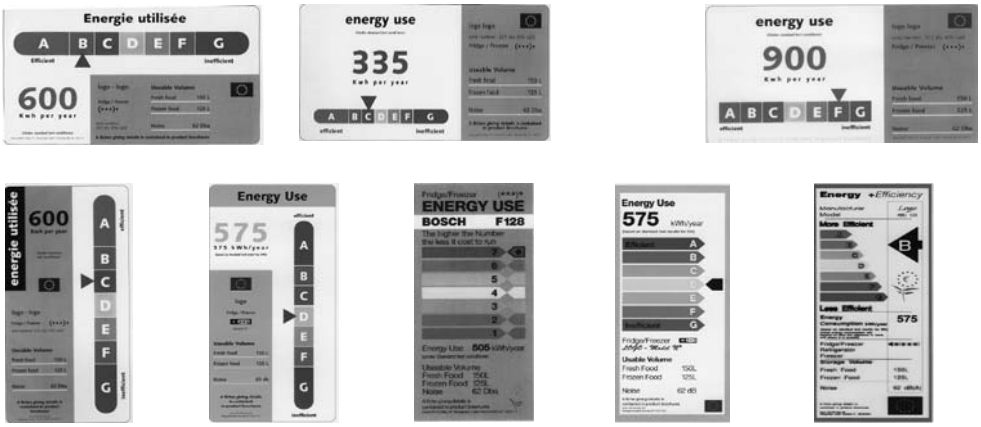

Figure 2. *Development of design concepts in Committee.*

- Utilise a two column format with the parameter on the left and the value in the adjacent column
- Block-off additional product information clearly differentiating it from the main energy information
- Be sure that the label primarily stimulates energy efficiency rather than communicates energy consumption. This information should be presented in an easy to process way
- Indicate relative energy consumption via simple categories

The use of categories in the EU label dates from this recommendation and the rationale that when consumers decide about the purchase of a domestic appliance they choose between options (price, height, etc.). If the number of options is large, consumers eliminate by aspects and judge options on the various dimensions in order of importance. If an option scores less than some minimum on an important dimension, it is excluded. Using categories helps simplify differentiation of energy performance and hence brings energy into the range of parameters to which this method can be applied.

From research to committee

The process of consumer research and expert consultation stopped at this stage of label development. Thereafter, the design was developed by an EC-led committee with support from a graphic design agency. It is not clear how the decision to use letters from A (more efficient) to G (less efficient) was taken, nor the use of a colour scale from green to red. Figure 2 shows the evolution of the committee produced designs. All include the experts' recommendations. The title of the label evolved before settling on just "Energy."

From A to A++: Did anyone think about consumers?

As with some other categorical energy labelling schemes the success of the EU label obliged a revision to take account of the fact that class A had become the dominant efficiency class. The EC ordered a technical study to consider the potential re-grading of the efficiency thresholds (Waide et al 2000), which established the need for a rapid revision of the refrigerator energy label including higher efficiency thresholds; however, no study was commissioned to consider what should happen to the appearance of a revised energy label. The technical study suggested that a simple rebasing of the existing A to G thresholds could be applied and that either a memory bar could be added to say what the label grade would have been under the older scale or that some other graphical aspect of the label could be changed (such as the background colour) to indicate that the new label was not directly comparable to the old. Perhaps under time pressure, the Commission and the Energy Labelling Regulatory Committee (composed of appointees from Member States) decided to adopt as a temporary solution a revised label with two new higher efficiency classes denoted A+ and A++. Unfortunately, this decision was taken without research on consumer comprehension of the proposed design change and with little consultation of market actors (e.g., retailers). In an attempt to fill the research gap quickly, a consumer focus group was organised in France by the French energy and environment agency ADEME to test the revised label candidates advanced by the industry association, CECED and the EC. Participants in this research misunderstood the A+, A++ concept and found it to be among the least favoured revision concepts. By contrast, participants understood and liked the idea of a changed background colour to denote a revision. They also indicated a preference for the middle class (D) to indicate market average efficiency and for the numeric efficiency index range to be indicated in parallel to the A to G scale. No testing was conducted of how consumers would react to the idea of entire classes at the low-end of the scale being empty of products, but the above suggests the reaction would not be positive.

Conclusions and impacts

The EU energy label has been an undeniable market transformation success and much of the credit must be attributed to its design. Market evaluations have shown a clear and strong evolution toward higher efficiency products since the label introduction, which contrasts favourably with the largely flat efficiency trends immediately prior to its announcement. Average energy efficiency is estimated to have improved by 37% for refrigerators, 21% for clothes-washers, and 35% for dishwashers since the label introduction at an

average rate of 4.0%, 3.7% or 6.5% per annum respectively (Waide 2004), though the same was not found for tumble driers and other products such as lamps have not been evaluated. Furthermore, there is clear evidence that the categorical label design has stimulated not only consumer demand but also manufacturers to develop products targeting specific higher efficiency thresholds both in advance of (i.e. in anticipation of) and in response to heightened consumer demand (Waide 2004). This demonstrates the clear value of using a categorical scale with thresholds that challenge manufacturers to develop more efficient products. Although the label clearly benefited from a mixture of direct consumer research and expert consultation, it is a pity that the latter designs, which included quite radical modifications, were not also researched. The use of a common label efficiency scale and format for all labelled products is also reported to have aided comprehension and "brand" recognition levels – the latter of which are said to be very high. Regrettably, there is no data available on the impact of the decision to add the A+ and A++ classes for refrigerators, but the little information there is suggests that consumers would have found a regrading of the existing A to G scale easier to comprehend.

A key suggestion that emerged from the expert consultation was the need to highlight important information either through font emboldening as well as to give thought to grouping or blocking off related information. This recommendation has since been borne out in many later label design research efforts including the US (Egan 2000a and b), India (Egan et al 2004) and China (Waide et al 2004). Though this may seem obvious it is a principle that has not been followed in the US-style continuous label which provides very little organisation of highlighting of key/related text.

INDIA

Since the passage of the 2002 Energy Conservation Act (EC Act) the Bureau of Energy Efficiency (BEE) is responsible for Indian energy efficiency policy. Prior to this a draft energy label was developed via market research split into three phases conducted sequentially from 1997 to 2000 (Dethman et al 2000) involving: 1) a survey of the attitudes and reactions towards the concept of labelling and energy efficiency amongst a sample of 1 833 major-appliance owners from 6 cities; 2) ten focus groups segregated by sex were conducted in three cities to design initial label formats using 11 designs based on international models pre-tested and a resulting in a final set of 17 potential labels; 3) expert consultation through a stakeholder focus group, that considered designs from the prior phases generated four label designs (Figure 3), which were subsequently tested in a simulated shopping environment with 673 appliance-owners in four cities.

- The initial findings of this research include (Dethman et al 2000, Deshpande 2001):
- Consumers preferred less technical terminology such as "power" or "current" to represent electricity and "units per day" over "kWh per day" to discuss quantities
- The concept of efficiency was not well understood or associated with appliances. The term power savings seemed to better convey the idea of effective energy use
- Approval rating for the concept of labelling was high with 70% saying they would use the label and 79% welcoming the label as a good idea
- Government endorsement enhanced the label's credibility
- While label designs using efficiency categories or continuous horizontal scales were tested and were well received, categorical approaches such as stars had better rates of comprehension in side-by-side tests of two labels with the same design
- A multi-phased approach where the questions addressed became progressively narrower until finally a final design emerged was seen as highly valuable

Follow-on Research Findings

Despite this excellent research, the label was not implemented due to a lack of legal authority. Subsequent to the EC Act, market research was undertaken to finalise the draft designs resolving issues that earlier research had left unclear such as: (a) the colour scheme of the label (experts suggested blue and green to connote the environment, but most of the consumer tests used red and yellow); (b) the best symbol between a hand holding money and hand holding lightning (experts thought the prior looked greedy and the latter aggressive); (c) reactions to the BEE logo not available at the time of the earlier research and (d) preferences between units consumed vs. rupees consumed. A secondary goal was to confirm the earlier results now three to four years old.

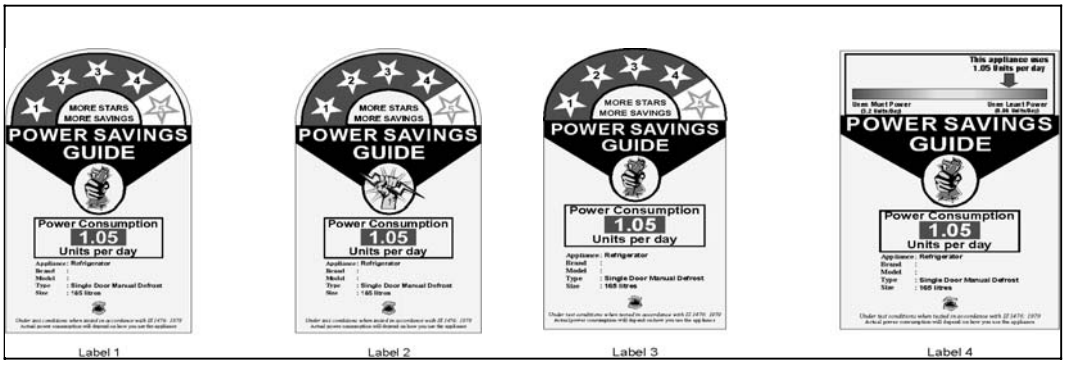

Figure 3. *The four trial labels presented to the expert panel.*

Sample labels and methodology can be found in Egan et. al (2004).

The qualitative research revealed that while most respondents expressed a spontaneous sense of curiosity when presented with the draft energy labels a potential initial barrier was that some consumers did not feel energy consumption was important at the time of purchase noting that if they thought about electricity consumption up-front then the conserving decision would be to not buy the product at all. Thus, labelling needed to be explained as showing meaningful differences in energy use that can impact bill savings.

Moving from the qualitative into the quantitative phase, the research focus shifted from reactions to labelling and the label designs in general to a more specific analysis of the label elements and their variation among the draft designs. According to the survey (Egan et al 2004):

- The unit and/or rupee consumption information presented in a central outlined box was the most noticed element on the label
- Although the BEE logo was not among the most noticed or liked elements, 90% of consumers and 90% of retailers felt it made the label look "more authentic and trustworthy"
- In an open-ended question that asked what consumers understood from the label, the vast majority made a correct general inference regarding the purpose of the label
- Just under one-tenth of respondents said that more stars mean more electricity consumption and therefore incorrectly concluded the star-rating worked in exactly the opposite direction intended
- The label showing unit per day and hand holding money symbol was preferred over the hand holding lightning and rupees per day
- In contrast to expert suggestion and the spontaneous mention by some consumers in the qualitative phase, the yellow and red colour combination was preferred over the green

Conclusions and impacts

This Indian research confirms the findings of many other comprehensive label design programs worldwide:

- the receptivity of consumers to the idea of energy labelling and its value as an information tool
- the stronger performance of categorical labels (e.g. stars) compared to labels with a continuous scale
- that a small portion of end-users, at least initially, conclude the opposite of the desired message that more stars mean more efficiency (Waide et al 2004, Egan 2000a & b)
- the importance of government endorsement to authenticate the logo
- the importance of understandable labels that avoid being overly technical
- the value of highlighting important information as the unit/rupee consumption information that was blocked-off in the centre of the label was also the most noticed element
- the importance of multi-method approaches to label design: in this case, as in research in the US (Egan 2000a) it was found that the qualitative methods revealed slightly different findings than quantitative

CHINA

The China National Institute for Standardization (CNIS) was the agency responsible for developing a mandatory energy information label for refrigerators in China and a comprehensive multi-method primary research project including: a consumer intercept survey; consumer focus groups; and semi-structured interviews with consumers, retailers, manufacturers, and policymakers was implemented. The research followed an iterative process and used a mixture of qualitative and quantitative methods to: (1) allow the maximum number of design concepts to be explored at each stage, and (2) progressively narrowing down the sets of viable designs by successive exclusion of the least successful concepts (Waide et al 2004).

Key findings of the qualitative phase, which tested more than 30 label trial designs, are summarised below by stakeholder. Additional details on the methodology and research design are in Waide et al (2004):

- The results of the retailer and manufacturer interviews were very consistent. Manufacturers were in favour of compulsory energy labelling as they felt it would provide a level basis for competition.
- Consumers appeared to have a very strong sense of the meaning of colour. Red was clearly associated with high energy consumption, green and blue with environmentally friendly/low energy consumption. Thus, designs using a green to red colour progression were strongly preferred to those without. In addition, green and blue emerged as the lead candidates for the label background colour.
- Consumers appeared to have a clear association that the letter A or the number 1 denotes the most energy efficient appliance (for labels using letters or numbers respectively).
- Labels using letters had the highest consumer comprehension (100% based on unprompted questionnaire results). Star labels were more likely to be misunderstood than those using letter or numbers with some consumers believing more stars meant more energy consumption. The comprehension of scales was aided by adding the Chinese character for 'class' next to each grading.
- The energy consumption value was an important parameter and most consumers wanted it emphasised over other product information (e.g., compartment sizes and noise). Consumers expressed a strong preference for daily not annual energy consumption apparently because prior manufacturer promotions had successfully conveyed the notion that an efficient refrigerator consumes less than 1 kWh per day.
- Consumers had a negative reaction to operating cost data with many saying they found it "too countryside," an ap-

Figure 4. *The final five label designs tested in the consumer survey.*

parent reference to this feature as lacking modernity and sophistication.

- Most consumers found designs with more than five efficiency categories confusing.
- The use of relatively large and bold characters or numbers was preferred to smaller ones.

Findings of the quantitative phase

The final round of research involved testing five optimised labels, Figure 4, via an extensive and quantitative consumer survey. Labels 1 and 2 are superficially identical, but 2 uses numbers to rate efficiency in place of letters. With the exception of label 5 all labels had the same background colour and all used the same fonts, information and descriptive text adjacent to the efficiency grades. Despite this, the labels vary by: numbers or letters, dials or stacked bars, vertical or horizontal layout.

To select the best candidate label design(s) with statistical precision, 1 250 people were surveyed in and around Beijing, Shanghai, Guangzhou and Wuhan. Almost 70% of respondents were able to correctly identify the most and least energy efficient model for all these optimised label executions. This comprehension rating is achieved amongst people who have never seen the label designs before thus an even higher level of comprehension could be expected following a promotional campaign and once a single design has been in place for a number of years. Despite the large sample there is no statistically significant trend in comprehension as a function of: income, education, and age although there is a significant influence according to the region and depending on whether the respondent had an agricultural or urban residency. Overall, the comprehension results of Labels 2 and 3 are significantly higher than for Label 4 while those of Label 2 are slightly higher than for Label 3. It should be stressed that these high comprehension scores are the product of the earlier qualitative and semi-quantitative design work, which had eliminated ineffective design concepts prior to the quantitative analysis stage.

Consumers were asked to rank on a 1 to 10 scale their ratings of each label execution for its ability to capture attention, its credibility, the appropriateness of the level of information, its ability to motivate the user to consider energy efficiency when making a purchase and its ease of comprehension. The questions were posed following exposure to individual label executions and reposed after exposure to all the labels. The individual exposures produced quite similar scores for all the labels with no label producing an average score of worse than 6.71 and none higher than 8.03. Again, these results suggested that all of the five final label designs worked well and indicated that lessons had been learned through the earlier research. By contrast, when all five labels were shown simultaneously, Label 4 scored far higher than the others with 35% of respondents choosing it as the one they would be most likely to read and 38% the most motivating, while only 15 to 16% give it the lowest rating for these parameters. It is very revealing that a similarly high share of respondents (35%) thought it was the easiest label to understand even though the true comprehension tests reported above found it was the least likely to be correctly understood. This demonstrates a very important factor in energy label design research: namely that consumer perceptions of which label is easiest to understand do not necessarily correlate with their actual levels of comprehension. In this case, it is quite possible that many of the factors they found appealing about the design were actually distracting them from the main message of the label. Based upon these results label 2 has been adopted as the mandatory energy label for refrigerators and with minor adaptation for room air conditioners.

The test of memory involved presenting respondents with five versions of the same label execution, each with a different energy efficiency grade and other lesser variations for the model name, manufacturer and daily energy consumption, in order to test the respondent's ability to correctly remember all of the five efficiency grades. The analysis showed there was no significant difference in 'recall' for the different label executions with all scores being between 26 and 30%.

Research to inform an a priori label decision

MEXICO

In 2002, CONAE the Mexican agency responsible for energy efficiency labelling since 1995, decided to undertake a revision of Mexico's label design. The decision was taken at the encouragement of manufacturers who suggested the label could be improved to give customers clearer signals on the efficiency level. CONAE undertook limited research prior to launch despite the fact that the new label would go into effect regardless of the results – as a check on its as-

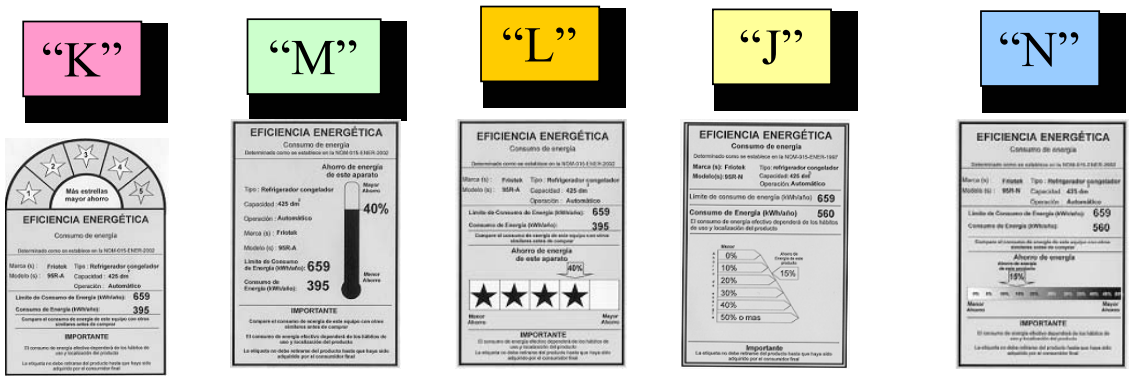

Figure 5. *The five trial designs tested in the consumer focus groups.*

sumption that the new design would be better than the old and effective for Mexican consumers. Thus, it was hoped the study would confirm an a priori decision about the new label design. Any comprehension or other issues the research might bring to light would inform outreach campaigns and potential future revisions. In late 2002 a total of six focus groups were conducted (two each in Mexico City, Monterrey and Guadalajara) with recent appliance purchasers and current shoppers. All the labels tested are shown in Figure 5.

The results showed that Label N (new to be implemented) and label J (old) scored higher in terms of credibility than did any of the alternate designs. Label N (new to be implemented) was the most credible. However, these two labels did very poorly when consumers were asked how appealing they found them. Labels K, L and M all scored more highly in terms of appeal. This relationship of an opposite tension between credibility and appeal has been a common finding in consumer research – with a strongly similar result found in the USA (Egan 2000a, b).

Participants were also asked how understandable they found the labels. Labels M and N (new to be implemented) scored relatively well on consumers perception of ease of understanding though not nearly as well as the leading candidate perceived ease of understanding--Label L. Overall, Label L was considered the best mix of appeal and perceived ease of understanding without diminishing the credibility of the label.

It should be pointed out the all the results of this study were qualitative and based on participants perception – i.e., not an empirical test of interpretation. This is particularly a concern regarding the issue of comprehension where international research has often shown a discrepancy between what people perceive as understandable and what they actually interpret as correct. This was a limitation that had to be accepted because of funding, but it confirms another truism of consumer research – that a multi-method analysis yields the most comprehensive results. The results also confirm that, as was the case in Europe, the US and India, participants tend to prefer categorical label designs to comparative designs.

SOUTH AFRICA

The South African Department of Mines and Energy launched energy appliance labelling based on a European

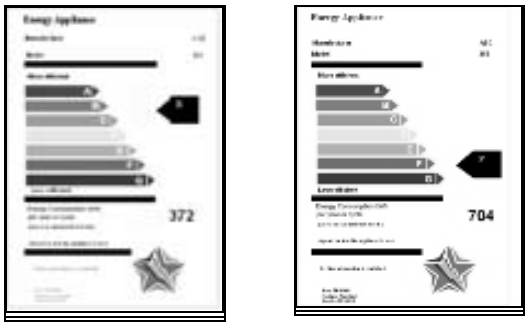

Figure 6. *The South African energy label as tested in the in-person survey.*

label design in May 2004. With this a priori decision taken, the program needed to know how to maximise consumers receptivity to the labels. Two consumer research tasks were implemented: a telephone survey and an in-person survey. The former focused on general questions regarding the appliance market, energy efficiency and the concept of appliance labelling. The latter focused on more specific questions – showing actual label designs to explore label comprehension, preferences and potential to motivate buying decisions. A side-by-side example of two models differing on energy was shown to facilitate a test of comprehension (Figure 6). See Surprenant et al (2005) for all methodological details.

Summary Label Findings

In general, the results of open-ended questions that asked conceptually about what might be included in an energy label (without actually showing label designs) support the main message contained on the South African A to G energy information labelling – particularly the concept of relative comparison of energy use. Further, they suggest that clear messages related to savings (particularly monetary) may have persuasive potential in communications. When consumers were asked how they would like the information on the labels to be presented, the most popular response was "graphically".

When shown the new energy labels in a side-by-side comparison of a relatively low- versus high-rating model, about half of the consumers could not determine which of the two

label showcards for two different appliance efficiencies used less electricity. One of the most significant differences between the labels was the least-noticed by all respondents (namely *"one has a B, the other has an F"*) though a majority (69%) did take note of that difference. Interestingly, the numerical difference between the two labels in annual energy consumption was slightly more noticed (by 79%) than the difference in letter grade. This is somewhat surprising as international research suggests consumers generally prefer and remember less technical information (e.g., letter grades) over technical data (such as annual energy consumption). The apparent struggle with the letter grading continued when respondents were asked what the difference in letter grading "told them." Only 41% of those who noticed the difference in letter grades in the first place (unaided) could provide a correct interpretation. Finally, about 40% were able to articulate a correct interpretation of the colour scale.

For the South African appliance label to be effective, consumers will need to be educated about how the label works—particularly the letter and colour scale. They may also need some encouragement regarding the labels credibility. Overall, the project demonstrates the importance of research to test label designs with end-users even if an a priori decision is taken on the label appearance (e.g., utilising a preferred international model) in order to: (1) identify key themes for promotion and outreach and (2) reveal local and culturally specific factors that can influence label comprehension. In addition, the finding that some significant proportion of consumers can, at least initially, have trouble taking away the main message from labels utilising a graded colour and/or letter scheme as used in the European label is consistent with research results in the US (Egan 2000a, b) but not that from China, the EU and Tunisia. Lastly, a stark contrast in comprehension rates can be seen in projects that design a label from first principles, as in China (where surveys showed ~70% comprehension of research optimised designs) and those that take make an a priori decision, as in this case (with comprehension rates of around 50%).

Research to help reach a final label design following a partial a priori label design decision

GHANA

The Ghana Ministry of Energy designated the Energy Foundation (EF) of Ghana as the official implementing agency for the Ministry's Energy Efficiency and Conservation Programme. As part of this responsibility, EF works with the Ghana Standards Board on the Ghana Electrical Appliance Labelling and Standards Programme (GEALSP). These stakeholders decided to implement energy labelling using a star-based label in Ghana. Taking this a priori decision as fact, the goal of this study was to develop an optimal star design for Ghanaian consumers.

A total of ten focus groups were conducted in four cities. Four sample labels were tested and for each of the four styles, one graph displaying a high-energy consuming air-conditioner model and one displaying a low energy-consuming air-conditioner model was shown side-by-side on a single showcard. These showcards were named M, O, P and T, Figure 7. Participants were asked to do two individual pen and paper exercises before the general discussions of the four showcards and one exercise afterwards. The first exercise that was done by participants tested their comprehension of the labels on the four showcards. For example on showcard M, were label M1 and M2. Respondents were required to choose from these two variants, M1 and M2, the one which, based on the information provided on the label, helps them best to decide on buying an appliance which will use less energy. With the second exercise, participants were asked to associate the four showcards to four selected statements to reveal their qualitative impressions of the labels. Lastly, participants were asked to give attribute ratings for the four showcards on a scale of 1 to 10. In general, the focus group tests applied addressed preferences towards and comprehension of the trial labels.

The results showed that:

Figure 7. *The trial energy labels tested in Ghana.*

Table 1: Rate of Correct/Incorrect Comprehension of Labels in Ghana.

	No. of correct choices	No. of incorrect choices
LABEL M	61	23
LABEL O	63	18
LABEL P	71	13
LABEL T	59	25

- none of the four labels tested were completely satisfactory to respondents in their existing format
- label T was preferred closely followed by label M. O and P were third and fourth respectively
- a few consumers thought more stars meant more energy consumption

However respondents had difficulty in comprehending Label T which had the highest number of incorrect choices among the four samples tested as shown in Table 1. Despite their actual comprehension problems, most respondents associated Label T with "best at communicating which appliance should be purchased assuming energy use was important".

Results showed that the ideal label in Ghana would have the following features: the number of stars on the label should be as in M/O but the size of stars should be made bigger; monthly operating cost should be included; the term EER should be explained; the outlined formatting of label T adopted along with a star arrangement as in P, M, O and a yellow background colour.

TUNISIA

The process and research behind the development of the new Tunisian energy label is described in paper 4,191 in these proceedings. The main findings were that it was possible to develop a relatively optimised label design loosely based upon the existing EU design, which was correctly interpreted by a remarkably high 86% of respondents who had never seen the label before.

Research to consider an amendment to an existing label

AUSTRALIA

The Australian government updated its 14-year-old appliance energy-labelling scheme, partly in response to the introduction of mandatory minimum energy performance standards for certain appliances that rendered the current efficiency rating system obsolete. It was the first time that a categorical energy label was revised and the efficiency categories "ratcheted" upward. As part of the labelling review, Australian policymakers wisely took the opportunity to assess not only the market and technical aspects necessary to redesign the label categories but to also consider consumer reactions to the label as a whole through consumer research. Two phases of market research were commissioned—the first to benchmark consumer understanding and acceptance of the current energy label and the second to devise a redesign that responded to consumer concerns (Artcraft Research 1998). In the first phase, a total of 7 focus groups were conducted. The response was clear and strong: the label in its current form was well-known, well liked and had a high degree of credibility. The star-rating scale was the most noticed element with the energy use figure trailing in importance to consumers. In fact, some consumers indicated they would prefer operating cost to kWh per year. The evaluation showed the marketing capital that a successful label can build with most participants affirming various label features such as the colour scheme and label name "Energy Rating." While respondents indicated that the main message of the star-rating was generally well understood, a common misunderstanding by nearly all consumers was that the label was a 5-star system when in fact the label was based on 6-stars. Even when shown an actual label many consumers could not see that there was room for a sixth star and it was concluded that an outline of all six stars with shading to represent the actual rating might aid appreciation of the six-star scale.

The problem faced by categorical labels over time of "bunching at the top" of the scale was tested with Australian consumers. Two options—adding more stars with products retaining their current rating versus retaining a 6-star scale but scaling current products down to allow for new, more efficient models—were tested. The scaling back option was clearly preferred in six of seven groups. "Many people felt leaving current products where they are was misleading in as much as it made it seem as if they were still efficient." Further, consumers wanted clear indications of the old label rating scheme in relation to the new label scheme presented on the retail floor. The new label is similar to the old label in colour and appearance, but the design is simplified and the font sizes and text positions are clearer to facilitate consumer understanding. There was also a conscious decision to visually separate the star rating at the top of the label (the part most commonly used by consumers) from the more technical data at the bottom of the label (energy, capacity, and so on) to make the label as friendly as possible.

UNITED STATES

The methodological approach and individual findings evaluation of the existing continuous-scale US label have been well documented (Egan 2000a, Egan 2000b). The preferred label design per this extensive, multi-phased and multi-method research is shown in Figure 8 and only the findings common to international lessons learned are presented here. They include:

The international trend toward categorical labels when tested side-by-side with continuous labels and American consumers clearly preferred a categorical style based on stars.

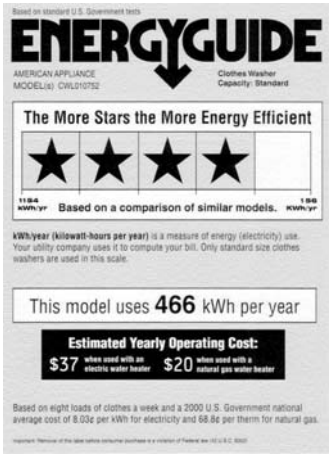

Figure 8. *The Preferred US label.*

- The need to highlight important information through font emboldening and further blocking off related information. In fact, a major problem with the US label is that consumers find it too cluttered, poorly organised and without relational grouping.
- The need to minimise technical terminology and use easy to understand and appealing visual images. In fact, a major problem with the US label is that consumers find it overly technical and graphically unappealing.

What is the value added by label research?

A good label design is a necessary, but insufficient condition for a labelling programme to achieve a significant market transformation impact. Having a good design, i.e. one which achieves high levels of: comprehension, motivation, appeal and credibility; does not of itself ensure that the labelling programme will be effective; however, the opposite is true such that if the label design is poor the labelling programme will be ineffective. The effectiveness of a label design can not be established by a committee of experts, but needs to be proven through valid market research. We briefly site a number of examples to support these findings:

- The Australian energy label has been found to be an effective vehicle to convey information on appliance efficiency in Australia and a rather similar label was found to be the best design of many in extensive consumer research carried out in India; however, when adapted to China the same label was found to have a very poor (near zero) comprehension levels due to different cultural interpretations of colours, dials, and stars.
- By contrast the colours used in the EU label (a progression from green to red) were found to be highly effective in China to reinforce the sense of the comparative efficiency scale.
- These two findings were not at all obvious prior to the conduct of market research and could not be assumed to be the case were the same labels tested in other cultures.
- The US EnergyGuide label had been in place for many years before research was conducted that demonstrated that a large percentage of consumers interpreted it in exactly the opposite manner to how it is intended (by interpreting annual operating cost information as annual operating cost savings). Despite this discovery the label has not been fundamentally altered to address this weakness.
- The Chinese, Tunisian and Indian research proves that a well designed label can be correctly interpreted by greater than 70% of the population, despite them never having seen the label design before. We would expect this share to rise after a period of familiarity with the labelling scheme and even more rapidly if the scheme is accompanied by a consumer awareness and retailer training campaign during its launch.
- Research from the EU and elsewhere has demonstrated how important it is to present the comparative energy performance of the appliance (i.e. the energy performance of the appliance in question compared to the range of energy use of appliances with the same functionality).
- The EU, Indian and Thai/US research has demonstrated the value of using discrete efficiency categories or classes rather than having a continuous scale. This finding has been supported by focus group results elsewhere too.
- The research results from Mexico and preliminary results from RSA indicate the risks from choosing a label without testing its performance first.
- Many label design research projects have demonstrated that information needs to be grouped, delineated and presented in a hierarchy of importance (e.g. by using font size and reading order to delineate importance). The corollary of this is that presenting too much information will reduce the labels effectiveness.
- If we sum all these conclusions together we would assert that label research can make the difference between having a label which has no positive impact on energy efficiency or could even be counterproductive and one which can accelerate the average sales-weighted efficiency of products on the market by several percentage points a year.

What are the best methods to adopt?

For countries designing a new energy label, it is clear that an iterative, multi-method research exercise is best. Ideally, the project should use a mixture of qualitative and quantitative research to reach an optimal and locally-relevant design. Beginning by examining a wide range of label design options based on international best practice, the best projects successively eliminate unsuccessful designs and refine promising ones. Examples of this kind of research process can be found in the label design exercises conducted in India, the USA and China and to some degree in the original work done in the EU and the limited label design research done in Tunisia.

Even if extensive multi-method research is conducted, considerable care has to be taken to avoid inadvertently introducing bias or limitations to the research. See Waide et al (2004) for a rigorous discussion of optimal methodology. The

introduction of bias/limitations can happen through: improper screening of focus group or survey respondents; inconsistent conduct of focus groups (through not preparing and consistently following a moderators guide); inconsistent or selective recording of responses; inadequate samples of trial label concepts; varying multiple label design factors simultaneously; not obeying a strict rotation of label design exposures (leading to systematic learning by doing); measuring inappropriate or incomplete parameters (e.g., perception of understanding instead of actual interpretation of take away messages). To this last point, research in the US, China and Ghana have demonstrated that the ease with which consumers state that labels can be comprehended often has little to do with their actual ability to be comprehended. Similarly, perceived comprehension of a label is often correlated with the perceived attractiveness of a label.

Conclusions

International research into energy information labels conducted to date has demonstrated the following general (i.e. universally applicable) findings:

- Label design by committee or policy and technical stakeholders rarely matches the needs of consumers as found in market research. As consumers are the intended end-users, new energy labels should always be designed through consumer-based market research
- Consumers, even those who have never been exposed to energy labels, generally think that comparative energy labelling is a good idea that would aid purchase decision making
- Labels that present the efficiency of the appliance on a comparative scale compared to other similar appliances are more easily understood and motivating than those that present technical information only
- Labels which present the comparative efficiency via discrete categories such as stars, letters or numbers are vastly more preferred and seem to be more effective than those which use a continuous scale. In part, this is because they are easier to remember when shopping for an appliance. In addition, the thresholds used in these labels can be highly motivating for both manufacturers and retailers
- In addition to a comparative efficiency scale it is helpful to highlight the primary energy consumption figure (such as the kWh used per period by a refrigerator, or the kWh per cycle used by a clothes washer)
- There can be strong connotations with colour and therefore it is helpful to exploit these to make the label more readily understandable and appealing
- Overloading the label with excessive or poorly organised information is distracting and limits both comprehension and engagement with the label. Careful blocking of related information and appropriate choices of fonts are helpful to make it clear to consumers which elements are most important and which only need to be addressed if further information is required
- The most appropriate design will depend upon local cultural factors and should be assessed by multi-method research. Often these cannot be foretold even by local policymakers as they lie outside of their expertise
- Each label design may have some limitations. For example, often a small portion of end-users at least initially conclude the opposite of the desired message that more stars mean more efficiency. And some studies have shown comprehension problems with letters based scales. These potential comprehension problems can and should be addressed over time through public education
- Government endorsement can often bring credibility to a label even in countries with historic bureaucratic problems. This is important as results confirm a tension between the credibility versus the appeal of label designs with technical looking labels make viewers feel confident in the labels authority, but detracting from making the label an eye-catching tool. A well-placed government endorsement can mitigate this impact
- The adoption of a well-known energy label design, even if it is successfully applied elsewhere, cannot be assumed to be effective in a new locale and hence this should, as a minimum, be confirmed through research before considering its adoption
- Proposed revisions to energy labels should be tested for effectiveness with key stakeholders (most importantly consumers) prior to adoption. Existing labels that do not undertake such evaluation risk losing hard won marketing leverage and brand equity
- Consumers often (but not always) express a preference for including operating costs on the label; however, no international labelling programme has resolved how to do this given that energy prices vary regionally and over time and given the high potential for confusion between operating costs and savings
- Policy-makers should aim to achieve 70%+ scores for consumer: comprehension, appeal, credibility, and motivational response from the label. Higher comprehension rates are typically found in labels designed from first principles than those designed through an a priori policymaker decision. Recall of the relative efficiency of competing products when shopping should also be measured
- If the effectiveness of an existing design has not been tested, it is appropriate to do so and to make amendments if the research demonstrates some significant weaknesses. Similarly, an existing design should not be modified without testing the effectiveness of the proposed change

Failure to follow these prescriptions would seem to seriously risk the integrity of the labelling programme and could risk minimising the energy saving and market transformation impact of the labelling scheme.

References

Artcraft Research. (1998) *Summary of Key Findings from the Second Qualitative Market Research Study Regarding Appliance Energy Rating Labels* and *Final Report on a Qualitative Market Research Study Regarding Appliance Energy Rating Labels*, Produced for the National Appliance and Equipment Energy Efficiency Committee, Australia.

Deshpande, V. (2001) "Labeling Design Effort in India: Lessons Learned." *Lessons Learned in Asia: Regional Symposium on Energy Efficiency Standards and Labeling*, Bangkok, 29-31 May.

Dethman L., Unninayar I. and Tribble, M. (2000) "Transforming Appliance Markets in India: Consumer Research Leads the Way." *ACEEE Summer Study*, Asilomar, CA.

Egan, C. (2000a). "Testing of International Labeling Approaches with US Consumers." EEDAL.

Egan, C. (2000b). "An Evaluation of the Federal Trade Commission's Energy Guide Appliance Label: An Interim Summary of Findings." ACEEE report # A003. Washington, DC.

Egan, C.; Dube, S. and Tathagat, T. (2004) "Introduction of Energy Labels: Assessment of Market and Consumer Research in India." *2004 DUE conference*, Capetown, South Africa

Gruppo Delta—Delta Group. (2002) "Qualitative Study to Explore Consumers Comprehension of the Mexican Energy Label." & "Project Delta—Energy Foundation: Final Report." Research International, Ltd.

Surprenant, L., Wegelin, J., Egan, C. and Mulholland, C. (2005) "Creating Relevant Appliance Labelling Campaigns Using National Telephone and Omnibus Surveys." To be presented at the DUE conference, Capetown, South Africa, March 29 to 31.

Surprenant, L., Wegelin, J. and Van Der Walt, T. (2005) "Results of the National Consumer Surveys Relevant to the Labelling Communications Campaigns: Final Report (DRAFT 2)", February.

University of Leiden (1991, 1992, 1993) *Energy Labelling: Report of two laboratory experiments*, Report E&M/R-91/26, *Energy information on domestic appliances: suggestions to optimize the Leiden vertical-label*, Report E&M/R-92/32 *Impact of energy information labels on consumer's choice of refrigerators*, Report E&M/R-93/36 Centre for Energy and Environmental Research.

Waide, P., ADEME, NOVEM, CECED, VhK, ENEA, TNO, Ecoles des Mines de Paris, Regent, Scholand, M. (2001) *Cold II: The Revision of Energy Labelling and Minimum Energy Efficiency Standards for Domestic Refrigeration Appliances*, DG-TREN, SAVE Contract XVII/4.1031/Z-98/269

Waide P., Egan C. and Minhong, J. (2004) "Findings of Energy Label Design Research in China" *ACEEE Summer Study.* Asilomar, CA.

Waide, P. (2004) *Energy labeling around the globe*, presented at *Energy Labels - a tool for energy agencies* 19 October, Brussels.