

Readability of product ingredient labels can be improved by simple means: an experimental study

Kerem Yazar¹, Gustaf Ö. Seimyr^{2,3}, Jiri A. Novak⁴, Ian R. White⁵ and Carola Lidén¹

¹Institute of Environmental Medicine, Karolinska Institutet, SE-171 77, Stockholm, Sweden, ²Department of Clinical Neuroscience, Karolinska Institutet, SE-171 77, Stockholm, Sweden, ³St. Erik Eye Hospital, 112 82 Stockholm, Sweden, ⁴Department of Design, Crafts and Arts, University College of Arts, Crafts and Design, SE-126 27 Stockholm, Sweden, and ⁵Department of Cutaneous Allergy, St John's Institute of Dermatology, St Thomas' Hospital, London, SE1 7EH, UK

doi:10.1111/cod.12274

Summary

Background. Ingredient labels on products used by consumers and workers every day, such as food, cosmetics, and detergents, can be difficult to read and understand.

Objective. To assess whether typographical design and ordering of ingredients can improve the readability of product ingredient labels.

Methods. The study subjects ($n = 16$) had to search for two target ingredients in 30 cosmetic product labels and three alternative formats of each. Outcome measures were completion time (reading speed), recognition rate, eye movements, task load and subjective rating when the reading of ingredient labels was assessed by video recording, an eye tracking device, and questionnaires.

Results. The completion time was significantly lower ($p < 0.001$) when subjects were reading all alternative formats than when they were reading the original. The recognition rate was generally high, and improved slightly with the alternative formats. The eye movement measures confirmed that the alternative formats were easier to read than the original product labels. Mental and physical demand and effort were significantly lower ($p < 0.036$) and experience rating was higher ($p < 0.042$) for the alternative formats. There were also differences between the alternative formats.

Conclusions. Simple adjustments in the design of product ingredient labels would significantly improve their readability, benefiting the many allergic individuals and others in their daily struggle to avoid harmful or unwanted exposure.

Key words: allergen avoidance; consumer products; contact allergy; cosmetics; food allergy; ingredient labelling; readability.

Product ingredient labelling is mandatory in the EU for food, cosmetics, and detergents, for good reasons;

exposure to specific ingredients may cause harm to individuals, and may even be fatal for those who suffer from some allergies. However, the usefulness of ingredient labels is lessened by the fact that they may be difficult to read, and this, together with often complex terminology, can make it difficult to avoid harmful or other unwanted exposures.

In a study from 2008, food-allergic subjects were instructed to purchase 15 specified foods that were potentially problematic from an allergy aspect (1). They were observed and interviewed in the shop. The subjects found the readability of the labels problematic, and complained about, for example, font size and colour contrast. The ability of patients with contact allergy to avoid exposure

Correspondence: Kerem Yazar, Institute of Environmental Medicine, Karolinska Institutet, Box 210, SE-171 77 Stockholm, Sweden. Tel: +46 8 52480042; Fax: +46 8 336 981. E-mail: Kerem.yazar@ki.se

Conflicts of interest: The authors have no relevant conflicts of interest to declare.

Funding: The Swedish Asthma and Allergy Association's Research Foundation for the submitted work, and general research programme support from the Swedish Research Council for Health, Working Life and Welfare (CL). The Sigvard and Marianne Bernadotte Research Foundation for Children's Eye Care (GS).

Accepted for publication 20 May 2014

to allergens in cosmetic products has been studied with qualitative research methods. The findings of Noiesen et al. (2, 3) have shown that many patients find it difficult or extremely difficult to read the ingredient labelling of cosmetic products. Noiesen et al. concluded that readability must be improved to facilitate allergen avoidance. In a questionnaire study in fragrance-allergic patients, 85% stated that clearer ingredient labelling would increase its benefit (4). According to the EU Cosmetics Regulation and its predecessor, the Cosmetics Directive, the information on the package should be written in 'indelible, easily legible and visible lettering' (5, 6). This is rather similar to the requirements for labelling of foods. A new EU regulation on food information for consumers will apply from 13 December 2014, and includes a requirement for a minimum font size (0.9 or 1.2 mm, depending on the package surface area) for mandatory information such as that in the ingredient label (7). Apart from this, there are no quantitative requirements in the EU regarding the typographical design of ingredient labels of products used by consumers.

The aim of reading a label is usually to obtain specific information, such as what a product contains or whether a product contains an allergen. The information is encoded in letters, numbers and words that can be expressed with varying legibility, depending on font, size, contrast, spacing, etc. Our eyes are remarkably efficient, even in poor visual conditions, but we can only see sharply in a very small portion of the visual field (six or seven letters at a time) (8). This means that our eyes move along the text while we are reading. For us to see clearly, the visual image needs to be stabilized on the retina in a 'fixation'. When reading, we move from fixation to fixation by using rapid eye movements called 'saccades'. Most saccades move forward to obtain more information, but some go back to previously read information to revalidate it. It is during the fixations that we decode letters and words, process the information obtained so far, and decide where to move the eyes next (9). Analysis of eye movement patterns can thus offer valuable insights into the reading process. Readability is commonly referred to as the ease with 'which the meaning of a text can be comprehended' (10), that is, how easy it was to solve the reading task. It is usually measured in terms of reading speed and comprehension, but, given the subjective nature of the task, it is also common to include additional inventories to better understand the reading situation.

The aim of this study was to assess how product ingredient labels are read by consumers, and to determine whether adjustments to the typographical design and different ordering of the ingredients on the labels could improve the readability. To do this, we designed an

experiment in which subjects read ingredient labels with different designs under controlled conditions while we recorded their performance.

Materials and Methods

A repeated-measurements experiment was designed in which each subject was instructed to find two target ingredients in 30 cosmetic products by reading the ingredient labels presented in four different formats (Fig. 1):

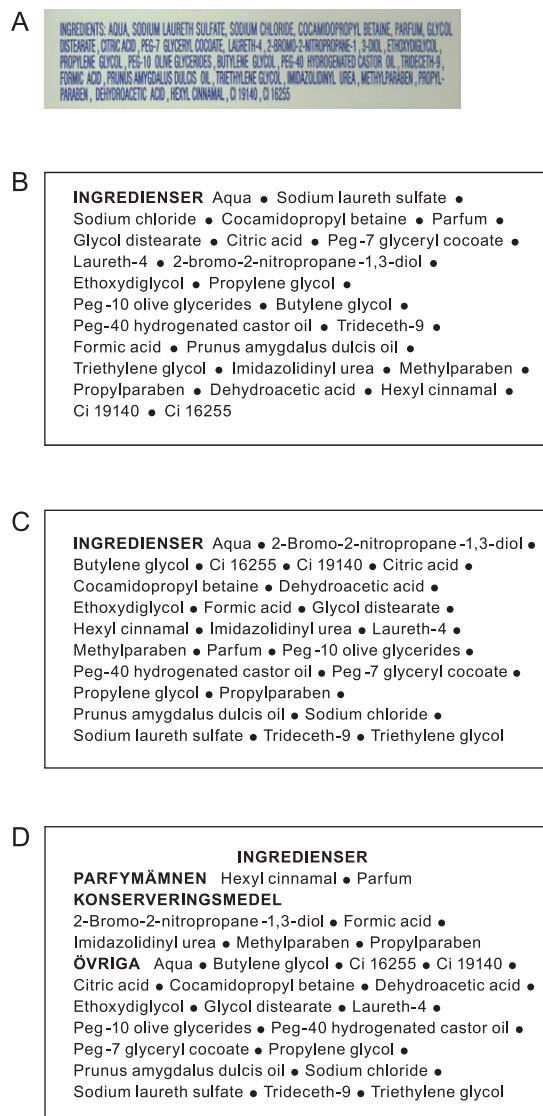


Fig. 1. Example of an ingredient label in each of the four formats (in scale 1:1). A = original typography on real products with ingredients ordered by concentration. B = enhanced typography with ingredients ordered by concentration. C = enhanced typography with ingredients ordered alphabetically. D = enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives.

- A Original typography on real products with ingredients ordered by concentration
- B Enhanced typography with ingredients ordered by concentration
- C Enhanced typography with ingredients ordered alphabetically
- D Enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives

Outcome measures

To evaluate how the formats affected readability, the reading speed was defined as the time required to determine whether or not a product contained the ingredients, and the recognition rate was defined as the ratio of correct responses. To determine how the formats affected the reading process, the eye movements of the subjects were recorded while they were reading, and the task load was rated by use of the NASA-TLX (Task Load Index) inventory (11) upon each format being finished. To determine which format the subjects preferred, a subjective experience and preference inventory was completed when the last format was finished. The null hypothesis of the experiment was no difference between the formats in completion time, recognition rate, eye movements, task load, experience, or preference.

Subjects

Sixteen subjects recruited by advertisement participated in the experiment (age 20–44 years, median age 23 years, eight females). The inclusion criteria were age 18–65 years, ability to read newspaper text without spectacles, and no heavy eye makeup, as these may interfere with the eye tracking device. Contact lenses were allowed. Two additional subjects were excluded, one because of a faulty audio recording, and one because of problems with the eye movement recordings.

The study adhered to the tenets of the Declaration of Helsinki, and was approved by the regional ethics review board in Stockholm, Sweden. All subjects gave written informed consent before taking part in the study.

Stimuli: ingredient labels

The 30 products used in the experiment were selected from a collection of 92 shampoos and 61 hair conditioners considered to be common on the Swedish market (12). The number of ingredients ranged between 6 and 47 (median 24). We wanted to reduce this range, as it was outside the scope of this study to discuss the composition of products. The products in the lower and upper quartiles,

in terms of number of ingredients, were excluded, leaving 80 products. In order to balance the sample in terms of readability, five colleagues of K.Y., who were not involved in the study, independently picked out the 15 products that they thought had the best readability and the 15 that they thought had the poorest readability. The 15 products picked by most as best and poor, respectively, were included in the experiment. The number of ingredients in the final sample ranged between 17 and 31 (median 24.5).

The typographical design of the ingredient labels on the selected original products was analysed, and areas of improvement were identified, on the basis of established knowledge of graphical design and readability. This formed the basis for the development of three alternative formats (Table 1).

The alternative formats were all typographically improved as described in Table 1; the difference among them was in how the ingredients were presented (Fig. 1). In alternative B, the ingredients were ordered by concentration (as in the original products). In alternative C, the ingredients were ordered alphabetically. In alternative D, the ingredients were ordered alphabetically and also had subheadings for fragrances and preservatives. These subheadings were chosen because these categories are among the most problematic groups of contact allergens in cosmetics. The alternatives were designed in a size that, with only minor rearrangement of the surrounding text, would fit on the selected products. Thirty ingredient labels, which were exact copies of the original products in terms of content, were produced in formats B, C, and D, respectively.

The alternative ingredient labels were printed on white paper (Colotech 100 g; Xerox, Norwalk, CT) and pasted on pieces of light grey cardboard (Folia 300 g; Folia Bringmann, Wendelstein, Germany) measuring $10 \times 15 \text{ cm}^2$. The ingredient labels of the original products were framed with cardboard (approximately 1.5-cm-wide edges) of the same light grey colour as for the alternatives (Folia 130 g; Folia Bringmann). This was done so that all ingredient labels would have a similar frame, and to help the subjects to focus on the ingredient labels during the experiment (Fig. 2).

Apparatus

The experiment took place in a dedicated light laboratory in a room of size $225 \times 320 \text{ cm}^2$, with a ceiling height of 290 cm. Two fluorescent tube armatures with microprism diffusers symmetrically placed over the subject produced a total of 500 lux with a colour temperature of 3300 K. The walls were covered with grey curtains in order to create a neutral and even environment covering the

Table 1. The main typographical shortcomings of the ingredient labels of a sample of cosmetic products (shampoos and hair conditioners, n = 30) purchased in Sweden, and the improvements that were applied when the alternative formats were designed

	Typographical shortcomings	Typographical improvements
Font size	Very small font size, often ≤ 4 points	Six points, corresponding to 1.5-mm upper case letters in Arial
Typeface	Typeface (i.e. bold, italic, condensed, extended) unsuitable for small font sizes. Serif typefaces not suited for the dominant printing technology (silk screen), in which contours tend to lose details and smudge	Arial normal, bold for headings and subheadings. San serif typeface suitable for this purpose, being well known and easily accessible
Alignment	To both left and right margins (justified), or centred; uneven gaps between words	Left-aligned. Eliminates major composition and typographic problems
Capital letters	All letters in upper case	Only first letter in upper case. Guides reader between ingredient names
Line spacing	Very narrow line spacing (leading)	Eight points (two points larger than typeface font size). Suitable leading is dependent on line length, typeface, font size, etc.
Contrast	Poor contrast between text and background, owing to colour combinations, glossy package material, and pictures/illustrations in the background	Black text on white (ideally off-white) background
Spacing and separation	None or unclear markings between ingredients; uneven and/or narrow spacing	Significant spacing between ingredients separated by a bullet (two points larger than the typeface font size, positioned on the baseline)
Splitting	Ingredient name on different lines	The whole ingredient name on the same line
Micro-typography	Poor micro-typography: individual letters and words too tightly or too loosely/unevenly kerned; compressed and/or extended typeface cuts; wrong letter spacing (tracking)	Optimized micro-typography: correct letter and word spacing; no compressed and/or extended typeface cuts



Fig. 2. Experimental setup and eye tracker. K.Y. searches for target ingredients on an original product.

subject's visual field. One researcher (K.Y.) performed the experiment in the laboratory with the subject. One other researcher (G.S.) monitored the recordings and logged the subject's progress and responses in the adjacent control room.

Eye movements were recorded with the Food and Drug Administration-approved Chronos Eye Tracking Device (C-ETD) (Chronos Vision GmbH, Berlin, Germany). It uses infrared video technology to record the eye movements.

The subject wears a helmet-like device that holds miniature video cameras, mirrors, and infrared light sources. For the experiment, the video cameras were configured to record eye movements at 100 Hz with a resolution of 200×150 pixels. An additional scene camera on the helmet recorded sound and the point of view of the subject at 30 Hz with a resolution of 640×480 pixels. With a scene camera, the eye tracker weighs ~ 500 g. No part of the helmet obscured the view of the products.

Procedure

Each subject was first informed about the experiment in writing, and was then asked to consent to participate. The eye tracker was introduced, and the subject was then comfortably seated in the light laboratory and introduced to the experimental task.

The subjects were instructed to search for the preservative cetrimonium chloride and the fragrance substance hexyl cinnamal in ingredient labels on original products and cards. They were informed that the ingredient labels could contain one, both or neither of the compounds. The subjects were asked to report the result orally after reading each label. Four products contained both of the ingredients, 14 contained one ingredient, and 12 contained neither of the ingredients. The ingredients

were declared on equal numbers of products regarded to have the best readability and to have the poorest readability. A slip of paper with the names of the two ingredients was placed in front of the subject throughout the experiment. The presentation order of the 30 products within each format (A–D) was randomized, and the order of trials with formats A–D was balanced over the experiment according to an orthogonal latin-square design to avoid test order effects.

Each trial with a new format (A, B, C, or D) began with a calibration of the eye tracker. The subjects were instructed to look sequentially at five dots in a cross-hair pattern on a piece of cardboard held at a fixed distance so that the visual angle between the dots was 6° . Then, the first ingredient label (original product or card) was handed to the subject. After reading the label and orally reporting the number of ingredients found, the subject handed it back and received the next, until all 30 products had been examined. The recording was then stopped, and the subject took off the helmet. The subject then proceeded with rating the task load by using the NASA-TLX inventory (11), in which mental, physical and temporal demands, and performance, effort and frustration levels, are rated on a continuous scale of 0–100 (low to high).

When all trials had been completed, subjects were shown the ingredient labels presented in all different formats, and asked to complete a subjective preference inventory. First, they selected which format they preferred, and then they were asked to rate each format in turn regarding readability, legibility, ease, speed and organization on a discrete scale of 1–5 (very poor

to very good). The order in which the formats were shown and rated for preference was randomized for each subject.

Analysis of eye movements

Eye movements were classified as transient, fixational, or saccadic, and loss of data was classified as either eye blinks or distortions. See File S1, including Tables S1–S3, for more details.

Statistical analysis

One-way repeated-measures ANOVAs with *post hoc* paired *t*-tests were used for parametric data (completion time and eye movement parameters). Friedman ANOVAs with *post hoc* Wilcoxon signed ranks tests were used for non-parametric data (recognition rate, task load, and experience). The preference was analysed with chi-square tests for ratios. Assumptions about normality were made *a priori*. All tests were two-tailed, with alpha set to 5% ($p = 0.05$). The *post hoc* tests were Holm–Bonferroni-adjusted for multiple comparisons. All statistical analyses were performed with ORIGIN PRO™ 8 software.

Results

The total duration of the experiment for each subject ($n = 16$) was ~ 1 hr, and ~ 30 min were spent on reading ingredient labels. The results for completion time, recognition rate, eye movements, task load and subjective preference are presented below (Figs. 3–5 and Table 2).

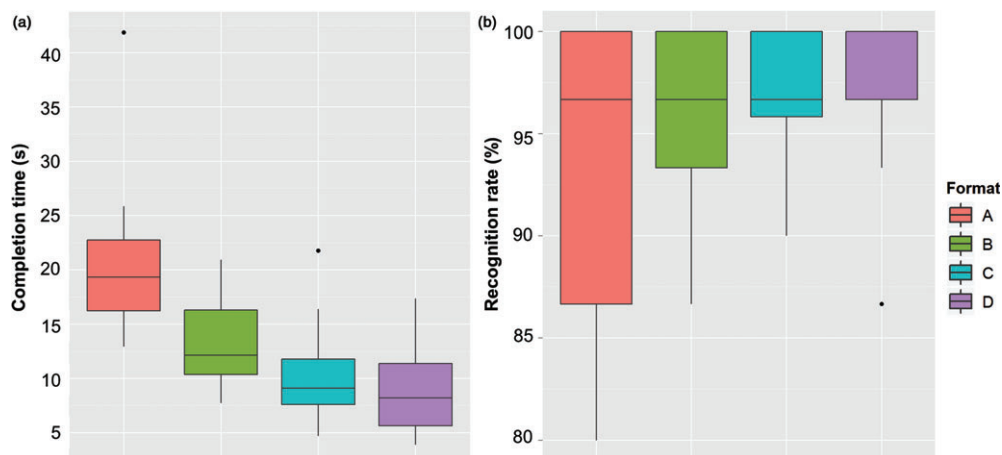


Fig. 3. Box-plot of completion time (a) and recognition rate (b) for the different formats. A = original typography on real products with ingredients ordered by concentration. B = enhanced typography with ingredients ordered by concentration. C = enhanced typography with ingredients ordered alphabetically. D = enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives.

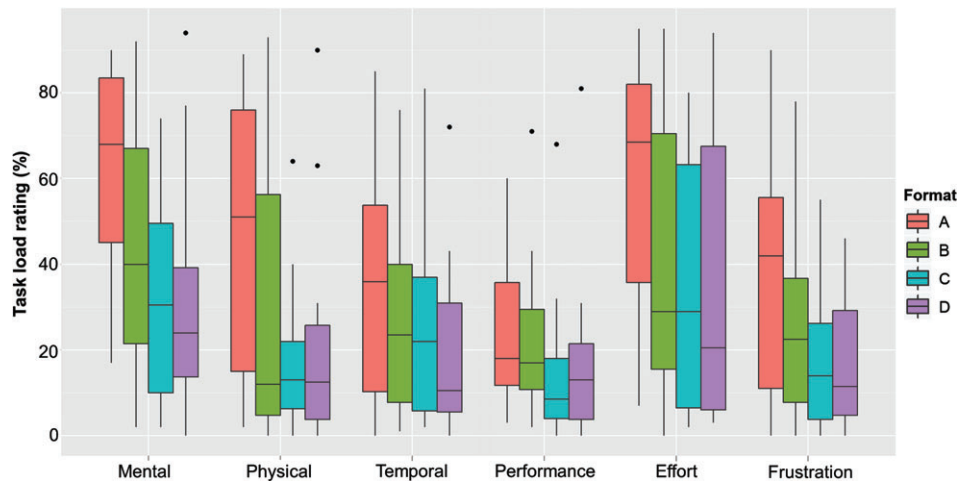


Fig. 4. Box-plot of task load ratings for the different formats. A = original typography on real products with ingredients ordered by concentration. B = enhanced typography with ingredients ordered by concentration. C = enhanced typography with ingredients ordered alphabetically. D = enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives.

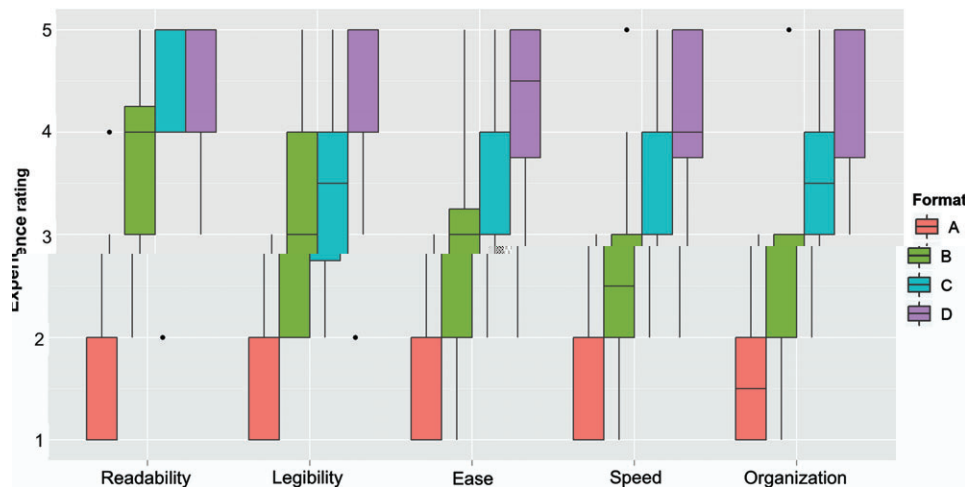


Fig. 5. Box-plot of experience ratings for the different formats. A = original typography on real products with ingredients ordered by concentration. B = enhanced typography with ingredients ordered by concentration. C = enhanced typography with ingredients ordered alphabetically. D = enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives.

Completion time

Completion time is a reflection of the reading speed, and was calculated as the time in seconds from the onset of reading a label to the oral response. The results were averaged over the 30 products for each format (Fig. 3a). There were significant differences between the formats ($F_{3,45} = 40.44, p < 0.001$). Pairwise comparisons showed that the completion time was significantly lower for formats B, C and D than for format A (the original product) ($t > 6.28, p < 0.001$). Completion time was also significantly lower for format C ($t = 2.68, p = 0.013$) and format D ($t = 3.75, p < 0.001$) than for format B.

Recognition rate

Recognition rate was calculated as the percentage of correct responses for which ingredients the products contained. There were no significant differences between the formats [$\chi^2(3) = 2.38, p = 0.50$]. The success rate was generally high, and was slightly better for formats B, C and D than for format A (Fig. 3b).

Eye movements

The eye movement recordings were analysed for each ingredient label from the onset of reading a label to the oral

Table 2. Key eye tracking results for different label formats; detailed results are given in File S1, Tables S1–S3

Measure	Label format and eye movements mean value (standard deviation)			
	A	B	C	D
Fixation count	59.69 (15.91)	42.23 (14.8)	33.94 (12.69)	30.99 (14.6)
Fixation duration (ms)	304.6 (48.96)	275.75 (27.84)	276.6 (37.33)	268.53 (33.84)
Saccade count	62.91 (17.48)	42.43 (14.22)	34.45 (11.91)	31.5 (14.64)

A, original typography on real products with ingredients ordered by concentration. B, enhanced typography with ingredients ordered by concentration. C, enhanced typography with ingredients ordered alphabetically. D, enhanced typography with ingredients ordered alphabetically and subheadings for fragrances and preservatives.

response regarding the number of ingredients found. Eye movements were classified as either saccades, fixations, transients, or distortions. Recordings with > 10% of data missing ($n = 32$) or an exceptionally high transient count (> 20 , $n = 61$) or distortion count (> 18 , $n = 5$) were excluded from further analysis. The results from the remaining samples ($n = 1822$) were averaged over the 30 products for each format.

The number of *fixations* per label differed significantly between the formats ($F_{3,45} = 29.65$, $p < 0.001$). Pairwise comparisons showed that formats B, C and D significantly reduced the number of fixations ($t > 5.21$, $p < 0.001$). The comparisons also showed that formats C and D significantly reduced the number of fixations ($t > 2.48$, $p < 0.017$) as compared with format B. The fixation durations differed significantly among the formats ($F_{3,45} = 10.68$, $p < 0.001$). Pairwise comparisons showed that all alternative formats significantly reduced the fixation duration ($t > 4.06$, $p < 0.013$). The drift of the fixations also differed significantly ($F_{3,45} = 7.4$, $p < 0.001$). Pairwise comparisons showed that all alternative formats significantly reduced the fixational drift ($t > 3.11$, $p < 0.003$) (Table 2).

The number of *saccades* per label differed significantly between the formats ($F_{3,45} = 32.85$, $p < 0.001$). *Post hoc* pairwise comparisons showed that formats B, C and D reduced the number of saccades ($t > 5.86$, $p < 0.001$) as compared with format A. The comparisons also showed that format D significantly reduced the number of saccades ($t = 3.13$, $p = 0.017$) as compared with format B. There were no significant differences in the amplitude or velocity of the saccades between the formats (Table 2).

Task load

Task load was enumerated as the number of millimetres to the left of the tick mark on the NASA-TLX scales. Friedman ANOVAS showed significant differences for all factors except temporal demand [$\chi^2(3) > 10.09$, $p < 0.018$] (Fig. 4). *Post hoc* analysis with paired Wilcoxon signed

rank tests and Holm–Bonferroni correction showed that the ratings of mental and physical demand, and effort, were significantly decreased for all alternative formats (B, C, and D) ($Z < -2.10$, $p < 0.036$). The alphabetical format (C) improved the performance rating significantly as compared with the original product labels ($Z < -2.11$, $p = 0.035$). The alphabetical format also reduced the mental demand significantly as compared with the typographical format ($Z = 1.97$, $p = 0.048$). The categorical format (D) reduced the frustration level significantly as compared with the original product labels ($Z < -2.31$, $p = 0.021$).

Subjective preference

Experience ratings were enumerated as a score between 1 (very poor) and 5 (very good) (Fig. 5). Friedman ANOVAS showed significant differences for all factors between all formats [$\chi^2(3) > 29.89$, $p < 0.001$]. *Post hoc* analysis with paired Wilcoxon signed rank tests and Holm–Bonferroni correction showed that the alternative formats increased the score significantly for all factors as compared with the original product labels ($Z < -2.03$, $p < 0.042$). The alphabetical format (C) increased the ratings for speed and structure significantly as compared with the typographical format (B) ($Z < -2.19$, $p < 0.029$). The categorical format (D) increased the ratings for legibility, ease, speed and structure significantly as compared with the typographical format ($Z < -2.32$, $p < 0.020$). It also increased the ratings for legibility and structure as compared with the typographical format ($Z < -2.03$, $p < 0.042$).

When the subjects were asked which format they preferred, none chose the original product label A. One preferred the typographical labels B; four preferred the alphabetical labels C; and the majority preferred the categorical labels D. A chi-square test of the preference indicated the difference in preference to be statistically significant [$\chi^2(3) > 9.58$, $p < 0.022$].

Discussion

Main findings

In this study, the readability of the ingredient labels of a sample of cosmetic products was compared with that of three alternative formats. To assess the readability, we used several measures – both subjective and objective – that are generally acknowledged and frequently used for this purpose. The readability of the alternative formats was significantly improved as compared with the original product labels. The alternative formats were designed on the basis of established knowledge of typographic rules and readability. The alternatives with the best design in terms of readability had typographical improvements, an alphabetical order of ingredients, and ingredients sorted into categories. The eye movement results show that the typographical improvement reduces the fixation times, and that alphabetical ordering and categorization of ingredients reduce the number of fixations needed to find an ingredient.

Strengths and limitations

We did not consider the problems regarding ingredient terminology in this study. The often long and difficult cosmetics ingredients INCI (International Nomenclature of Cosmetic Ingredients) is the legal requirement in the EU; names have been shown to be a significant obstacle when trying to avoid allergen exposures (3). Similar problems have been discussed concerning food labelling (13). This question was considered too complex to be handled within the scope of the present project. Notwithstanding this limitation, the improvements in readability were large, and unambiguously shown by several objective and subjective measures. The improvements were achieved by applying broadly recognized typographical rules and principles to the original text. This proves that it is relatively easy to produce more accessible product information, although the ingredient names may be difficult. The experimental design in combination with objective measurements was cost-effective and suitable for the purpose of the study.

Other studies

There is, to our knowledge, no prior research on how ingredient labels are read by consumers, or on how they should be designed to improve readability. However, there are a number of studies on nutrient labelling (14). The aim has typically been to assess the understanding and use of various labelling formats with largely subjective measures, but some studies have also used eye tracking to obtain objective data (15–17). The results suggest that enhanced legibility and highlighting of specific

nutrients can improve readability. Although ingredient and nutrition labelling do not serve the same purpose, they still share many important properties. Their inclusion is typically regulated, and the space allowed for them is limited. They are attached to consumer products with varying sizes, shapes, and colours. Both carry important information for the consumer presented under similar, and often challenging, conditions for reading.

Possible implications

In the new EU regulation on food information for consumers, it is stated that an approach should be developed that considers all aspects related to legibility, including font, colour, and contrast (see preamble point 26) (7). In addition, the European Commission should establish rules for legibility (§13, point 4). Several features of the alternative formats in our study could (rather easily) be introduced as requirements in various EU regulations concerning cosmetics, food, household products, etc. It should also be said that industry could introduce several of these requirements voluntarily now, even though they are not yet mandatory.

We wish to address some of the most important shortcomings and suggest improvements:

Font size. The font size was, in some cases, so small that even people with normal vision experienced problems in reading the original ingredient labels. A minimum font size requirement would be an important step towards improved readability. Such requirements are already specified in the United States for labelling of cosmetic ingredients, and in the EU regulation on food information (which will apply from December 2014) (7, 18). We suggest that regulators also follow these examples regarding other product types, such as cosmetics and other chemicals.

Poor contrast. Poor contrast between text and background impairs the readability severely. This problem could be easily eliminated by the use of standardized labels with black text on a white background. It may be argued that such a requirement is unfavourable from a product design aspect. However, it should be recognized that the primary function of ingredient labelling is to inform consumers, being a tool for safety and health protection and essential for allergic individuals. Thus, aesthetic aspects should not take precedence.

Alphabetical order. According to the current EU regulations on food and cosmetics, ingredients should be listed in descending order by weight. It was instructive to see

the substantial positive effect that was achieved by merely alphabetically sorting the ingredients. All measures in our study showed this. To the patient, consumer or worker who needs to avoid exposures to a specific ingredient, there is no meaning to ingredients being listed by weight, as required now. We thus suggest that alphabetical order should be used for ingredient labelling.

Other features. Additional features that would be beneficial and that appear to be easy to implement are clear separation between ingredients, left alignment of the text, and typeface and line spacing suitable for the font size.

Future research needs

Our study had its starting point in considering cosmetic ingredient labels in which complex INCI names may be difficult to understand. It would be of considerable interest to study how the terminology of ingredients could be made simpler and more accessible to consumers.

References

- Cornelisse-Vermaat J R, Voordouw J, Yiakoumaki V, Theodoridis G, Frewer L J. Food-allergic consumers' labelling preferences: a cross-cultural comparison. *Eur J Public Health* 2008; **18**: 115–120.
- Noiesen E, Larsen K, Agner T. Compliance in contact allergy with focus on cosmetic labelling: a qualitative research project. *Contact Dermatitis* 2004; **51**: 189–195.
- Noiesen E, Munk M D, Larsen K, Johansen J D, Agner T. Difficulties in avoiding exposure to allergens in cosmetics. *Contact Dermatitis* 2007; **57**: 105–109.
- Lysdal S H, Johansen J D. Fragrance contact allergic patients: strategies for use of cosmetic products and perceived impact on life situation. *Contact Dermatitis* 2009; **61**: 320–324.
- Council of the European Union. Council Directive of 27 July 1976 on the approximation of the laws of the member states relating to cosmetic products (76/768/EEC). Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1976L0768:20090507:EN:PDF> (last accessed 10 June 2014).
- European Parliament, Council of the European Union. Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R1223-20140401&from=EN> (last accessed 10 June 2014).
- European Parliament, Council of the European Union. Regulation (EU) No. 1169/2011 of the European Parliament and if the Council of 25 October 2011 on the provision of food information to consumers, amending Regulation (EC) No. 1924/2006 and (EC) No. 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No. 608/2004. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1402415318268&uri=CELEX:02011R1169-20140219> (last accessed 10 June 2014).
- Robeck M C, Wallace R R. *The Psychology of Reading: An Interdisciplinary Approach*, 2nd edition: Hillsdale, NJ, L. Erlbaum Associates, 1990.
- Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 1998; **124**: 372–422.
- Mills C B, Weldon L J. Reading text from computer screens. *Comput Surv* 1987; **19**: 329–358.
- Hart S, Staveland L. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: *Human Mental Workload*, Hancock P, Meshkati N (eds): North-Holland, Elsevier Science Publishers, 1988.
- Yazar K, Johnsson S, Lind M L, Boman A, Lidén C. Preservatives and fragrances in selected consumer-available cosmetics and detergents. *Contact Dermatitis* 2011; **64**: 265–272.
- Joshi P, Mofidi S, Sicherer S H. Interpretation of commercial food ingredient labels by parents of food-allergic children. *J Allergy Clin Immunol* 2002; **109**: 1019–1021.
- Cowburn G, Stockley L. Consumer understanding and use of nutrition labelling: a systematic review. *Public Health Nutr* 2005; **8**: 21–28.
- Ares G, Gimenez A, Bruzzone F, Vidal L, Antunez L, Maiche A. Consumer visual processing of food labels: results from an eye-tracking study. *J Sens Stud* 2013; **28**: 138–153.
- Antunez L, Vidal L, Sapolinski A, Gimenez A, Maiche A, Ares G. How do design features influence consumer attention when looking for nutritional information on food labels? Results from an eye-tracking study on pan bread labels. *Int J Food Sci Nutr* 2013; **64**: 515–527.
- Jones G, Richardson M. An objective examination of consumer perception of nutrition information based on healthiness ratings and eye movements. *Public Health Nutr* 2007; **10**: 238–244.
- Food and Drug Administration. 21 C.F.R. § 701.3.

Conclusion

Our study shows that rather simple adjustments in the design of consumer product ingredient labels would improve their readability and clarity significantly; this would be of great benefit to the many allergic individuals and others in their daily struggle to avoid harmful or unwanted exposures and negative health effects.

Qualitative and quantitative requirements regarding the typographical design of ingredient labels should be introduced in relevant EU regulations.

Acknowledgements

We thank the study subjects for their participation.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

File S1. Eye movement analysis.