Journal of Engineering Design

ISSN: 0954-4828 (Print) 1466-1837 (Online) Journal homepage: https://www.tandfonline.com/loi/cjen20

Beyond usability: designing for consumers' product experience using the Rasch model

Fabio R. Camargo & Brian Henson

To cite this article: Fabio R. Camargo & Brian Henson (2015) Beyond usability: designing for consumers' product experience using the Rasch model, Journal of Engineering Design, 26:4-6, 121-139, DOI: 10.1080/09544828.2015.1034254

To link to this article: https://doi.org/10.1080/09544828.2015.1034254

Published online: 14 May 2015.

Submit your article to this journal

Article views: 432

View related articles

View Crossmark data

Citing articles: 9 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cjen20

Journal of Engineering Design, 2015
Vol. 26, Nos. 4–6, 121–139, http://dx.doi.org/10.1080/09544828.2015.1034254

Beyond usability: designing for consumers' product experience using the Rasch model

Fabio R. Camargo and Brian Henson*

School of Mechanical Engineering, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK

(*Received 31 August 2014; accepted 23 March 2015*)

Many manufacturers wish to relate the physical properties of their products to customers' affective responses so that they can improve the design of their products. Rasch measurement theory is a novel approach in this context that could offer business benefits over other approaches. The aim of the research was to determine whether Rasch measurement theory can support the construction of a scale to measure the affective impression of a moisturiser cream elicited by the compliance of the moisturiser's packaging. The compliances of packaging used to develop a scale for the impression of the contents of the packaging in previous research were measured. New packaging prototypes were manufactured with a range of compliances to determine whether affective responses to them could be predicted. A new sample of 67 participants rated the prototypes against Likert statements on a five-point scale. The statements were derived from descriptions of the users' product experience. The scale was recalibrated using a multi-facet Rasch model. It was found that the affective responses to the prototypes fell within the range predicted by the original scale. The results demonstrate that Rasch measurement theory can be used to achieve the objective correspondence between affective responses and the physical properties of a product, and the information can be used in new product development.

Keywords: experience design; human factors engineering; affective engineering; kansei engineering; Rasch measurement theory

1. Introduction

This paper is about measuring people's emotional responses, sentiments or attitudes to products, so that the relationships between the formulation, physical properties or features of a product and the responses can be modelled and used to inform the product's design. The scope of these responses could include, for example, whether a motor vehicle is pleasant to drive, and aspects of the product experience, such as whether a razor feels as though it is the 'best a man can get',[1] or whether some laundry detergent 'gets to the heart of the wash',[2] and affords the launderer the satisfaction and fulfilment of caring for his/her family. These emotions and attitudes reflect latent variables or constructs which cannot be directly observed, and which cannot be measured as easily as other objective aspects of the product. They reflect aspects of the total product experience which go beyond the products' functionality or usability, and are associated with the emotional benefits provided by using the product. Consequently they are influenced by culture, personality,

*Corresponding author. Email: b.henson@leeds.ac.uk

© 2015 Taylor & Francis

experience and by the socio-economic status of the consumer. These latent constructs are important, because they are often the principal discriminator between competing products (Childs et al. 2006). In this paper, the emotions and attitudes about products are referred to as the affect, and their expression by people is referred to as their affective response.

The aim of the experiment reported in this paper was to determine the correspondence between the physical properties of a product and affective responses using Rasch measurement theory. Approaches to quantification of people's affective responses to products include counting, such as counting the number of consumers who say that they would recommend a product to a friend, and ranking, such as when asking consumers to place a range of products in order of their preference. However, such techniques do not provide interval-level measurement. Rating scales can be used on which people are asked to respond to statements (as in Likert scales (1932)) or contrasting adjectives (as in the semantic differential scale (Osgood, Suci, and Tannenbaum 1957)) on, typically, five- or seven-point scales. The values assigned to responses in each category are then taken as some sort of quantification of an underlying construct. These scales do not provide interval-level measures (Stevens 1946), although they are often used as though they do. Ratings are sometimes analysed using factor analytic techniques such as factor analysis, principal components analysis (PCA) (Snider and Osgood 1969) and multi-dimensional scaling (Shiffman, Reynolds, and Young 1981) to construct semantic spaces in which to visualise the responses.

A disadvantage of rating scales is that they rely on the honesty and awareness of the respondent for their validity. Techniques used to quantify affective responses using objectively manifested responses include analysis of facial expressions (Zeng et al. 2009) and electroencephalography (Schaaff 2009). The challenge with these techniques is that the features taken as a proxy for the underlying construct need to be identified and signals from different sources meaningfully combined (Gunes, Piccardi, and Pantic 2008).

There are a number of techniques from the domain of psychophysics for measuring sensory aspects of products which are beyond the scope of this paper (see, for example, Kingdom and Prins 2009), and the concept of relating sensory and affective responses to product design is well established in the domain of food science, many of the principles for which were established in the 1950s and 1960s (Drake et al. 2009). Extension of the concept to the quantification of affective responses and their relationship with product properties is not so well established in product design, and many researchers cite Nagamachi's kansei engineering (1995) as the initiation of the concept in human factors engineering. In Nagamachi's approach, based on Osgood's semantic differential technique (Osgood, Suci, and Tannenbaum 1957), a large number of people (typically many dozens) are asked to rate a product against a number of adjective pairs (antonyms or declarative opposites) on multi-point scales. The responses to the questionnaires are turned into a quantification of affective response using PCA to create a semantic space against which the physical properties of the products can be correlated. In the authors' view, many published kansei engineering studies have adapted and applied Osgood's method poorly by, for example, using experiment design with inadequate samples, and do little to promote rigour in measurement of affective responses.

From a commercial perspective, the principal problem with this approach is that it is expensive. If a manufacturer measures the affective responses to the use of a new fabric detergent using a large sample of consumers, changes to the detergent's formulation have to be assessed by repeating the entire study. Furthermore, the approach often does not match the industrial product development process. Manufacturers usually wish to know about specific constructs identified from market research which are associated with particular product experiences, trends or brand values. However, PCA lets constructs emerge empirically from the data, and it is often difficult to relate the resulting semantic space to the constructs the manufacturer wishes to know about. For example, in a study looking for fresh and real qualities of packaging, principal components loading on the adjectives 'appealing' and 'genuine' emerged from the data (Childs et al. 2006).

From a technical perspective, if the process is not implemented with care, inaccuracies can be introduced. These include treating the scales as interval, when they are at best ordinal (Wright and Linacre 1989; Stevens 1946); the assumption that adjective pairs plotted on the semantic space are equidistant from the neutral point (Heise 1969) and the assumption that common word pairs are antonyms (Mordkoff 1963). Unwanted sources of variance, such as imprecision in reliability, factor scores and group means, also cause inaccuracies (Borsboom 2006). While it is not entirely clear to what extent the inaccuracies affect quantification, it is widely accepted that they result in a non-linear correspondence between people's affective ratings and design elements. Asking people to make their ratings on continuous lines (the visual analogue scale) instead of according to categories does not make the response linear (Thomeé et al. 1995). The use of neural networks (Ishihara et al. 1995), fuzzy logic (Hotta and Hagiwara 2005) and rough set theory (Nagamachi 2008) has been proposed to overcome problems of non-linearity in data. These approaches, however, lack metrological rules such as traceability and variance control (Rossi 2007), and therefore remain ones of analysis of a particular sample without allowing results to be compared. Because of the difficulty in comparing outcomes of studies, many of the things manufacturers would like to know are difficult to quantify, such as comparisons of the preferences of different market segments or evaluations of products over time.

The Rasch measurement model has been proposed to address these problems. The Rasch approach uses probabilistic models of peoples' responses rather than descriptive statistics, was pioneered in education and is now widely established in medicine and psychological testing (Rasch 1960 [1980]; Embretson and Reise 2000). Data from questionnaires which fit the model are transformed into linear measurement (Rasch 1960 [1980]; Luce and Tukey 1964; Karabatos 2001). The measurement has a quantitative structure that demonstrates additivity and invariant comparisons (Andrich 1988). Demonstration of invariant comparisons is a property of the model which means that comparisons between individuals are independent of the statements on the questionnaire and the particular products in the study; and, conversely, that comparisons between products are independent of the individuals and statements. Rasch theory is related to item response theory (Embretson and Reise 2000), although many Rasch practitioners and researchers distance themselves from the association because, they assert, the Rasch model uniquely addresses fundamental issues of measurement. Despite its widespread use in other domains, Rasch theory has not yet been widely used in measurement of product evaluation, although it has found some application in marketing (Salzberger 2009).

The Rasch approach might offer advantages when attempting to identify the physical properties or features of products that elicit particular emotional or affective reactions from individuals. From a commercial point of view, the principal benefits of using the Rasch approach perhaps lies in the ability to calibrate questionnaires. Once calibrated, it should be possible to administer questionnaires to small samples without loss of reliability, and develop computer adaptive implementations of questionnaires. These should substantially reduce the costs of measuring consumers' affective responses to products.

In this paper, research is reported in which it is demonstrated that Rasch measurement theory can be used to identify an ideal range of compliance values for the packaging for moisturiser creams. The research is novel because it demonstrates that a scale can be used to relate affective responses to physical properties of a product, and consequently be used to design new products. A brief description of the Rasch model used in the research is described in Section 2. It is included for readers who might wish to refer to sources on the mathematics of the model, and an appreciation of the contents of Section 2 is not necessary for an understanding of the research. Previous research that establishes the theoretical underpinning for the work is outlined in Section 3. Method, results and discussion are included as Sections 4, 5 and 6, respectively.

2. Theory

Rasch (1960 [1980]) expressed his original model in terms of the correctness of an individual's response to the difficulty of reading assessment. Here, the language is transposed to that of product evaluation, in which an item refers to a response to a statement or adjective pair on a questionnaire. Thus Rasch's model is the conditional probability that a person will endorse a statement (or item) against a dichotomous scale (e.g. yes or no) as a logistic function of the difference between the person's inclination to endorse the statement (β) and how easy it is to endorse the statement (δ) on a linear scale (Equation (1)):

$$\Pr\{X = 1|\beta, \delta\} = \frac{\mathrm{e}^{(\beta-\delta)}}{1 + \mathrm{e}^{(\beta-\delta)}}, \tag{1}$$

where $\Pr\{X = 1|\beta, \delta\}$ is the probability that the person will endorse the statement.

The basic model was evolved to take into account polytomous responses, such as category scales (Rasch 1961; Andrich 1978; Masters 1982), and Linacre extended the model to include other facets of the measurement that affects the scores (Linacre 1989, 2002). Linacre's multi-facet model has been adapted for use in product evaluation, in which many people are evaluating many products against multi-categorical items (Camargo 2013). In this model, the facet 'stimulus' is associated with products or physical elements of products (Equation (2)):

$$\Pr\{X_{nis} = x_{nis}|\beta, \delta, \zeta\} = \frac{\exp\left[\sum_{k=0}^{x} (\beta_n - \delta_i + \zeta_s - \tau_{isk})\right]}{\sum_{k=0}^{m} \exp\left[\sum_{k=0}^{x} (\beta_n - \delta_i + \zeta_s - \tau_{isk})\right]}, \tag{2}$$

given that the denominator is a normalising factor, and $\Pr\{X_{nis} = x_{nis}|\beta, \delta, \zeta\}$ is the probability of a respondent n giving a rating of $k, k \in (1, \ldots, m)$, on item i for stimulus s; β_n is the inclination of a person to endorse the statement i for stimulus s; ζ_s is the level of the attribute fulfilment of the stimulus s; δ_i is the difficulty of endorsement of statement i and τ_{isk} is the threshold parameter given a rating of k on the statement i for stimulus s.

An example of how the model is used for analysis of affective responses is outlined in Section 4. Unlike in statistical approaches which model or describe the data, the approach in Rasch theory is to use computer-based iteration to determine whether the data fit the model.

3. Previous work

Previous research has established that Rasch measurement theory can be used to develop scales of people's affective responses to products (Camargo and Henson 2011), that scales support invariant comparisons of different product characteristics (Camargo and Henson 2012) and that scales can be stable when new items or statements are added (Camargo and Henson 2013). The ability to calibrate and compare the scales has been used to demonstrate that affective responses to touching fabrics for vehicle seats depend on context (Henson and Camargo 2014).

In a first attempt to use Rasch measurement theory to develop a scale for affective responses, Camargo and Henson (2011) asked 306 participants to rate their endorsement of 24 Likert statements related to the construct of *specialness* for four pieces of confectionery. The confectionery was Caramel® and Milky Way® from a Mars Celebrations® assortment, Ferrero Rocher® and Lindor®. The statements had been determined through consumer research by an international confectionery company, and many of them were based on verbatim comments used by consumers during focus groups to express the emotional benefits of consuming chocolate. Example statements are, 'A box of these chocolates would make a lovely romantic gift' and 'I would keep chocolates like this one for myself'. Likert statements are preferred because, whereas PCA creates a multi-dimensional construct space from the responses to the adjectives, the Rasch approach deals with unidimensional constructs, and it is difficult to identify a sufficient number of adjectives to define a unidimensional construct of *specialness*. Furthermore, it is relatively straightforward to generate Likert statements that capture nuances of the persons' usage scenarios and product interaction that demonstrate some face validity for measurement of the constructs of interest.

The responses to the four pieces of confectionery were initially analysed separately, to identify a calibrated measure of *specialness* for each confectionery, which revealed differences in response patterns. Statements appeared at different locations on the scales' continuum for each confectionery, preventing direct comparison between the scales and establishing the need to include the stimulus as an independent parameter in a multi-facet model. A further calibration step using the multi-facet model (Equation (2)) established a unified scale, for which 12 statements fitted the model for all confectionery. The rejected statements may have been redundant due to high correlations between them, and/or ambiguous. Some statements were rejected because they were poorly targeted, meaning that they were above or below the range of measurement captured by the scale, such as when hardly anyone or almost everyone endorsed a statement. The responses of around 10% of the participants were removed for calibration of the statements. The work demonstrated the establishment of linear scales for the measurement of the specialness of confectionery which allowed comparisons between products, individual persons, and between statements.

For the use of the proposed multi-facet model (Equation (2)), it remained to be demonstrated that the locations of statements were stable on the scale and do not change when new calibrated items are added. That the position of statements should not change as new statements are added to the questionnaire is a prerequisite for the implementation of computerised adaptive testing (CAT). In product design, CAT offers the potential to reduce costs in product evaluation because the time needed for data collection could be substantially reduced. CAT is concerned with establishing a sequence of statements selected by computer such that if a respondent endorses a statement, a more challenging statement for endorsement is presented in the sequence, and contrariwise if the statement is not endorsed. This converges into a sequence of statements bracketing the person's endorsement level. Consequently, each person does not answer all statements in the bank of statements, only a subset bracketing the threshold of endorsement. The technique is well developed in education (Wright and Bell 1984). As new statements are added to the item bank, the positions of the existing statements on the scale should not change.

Camargo and Henson (2013) have demonstrated the stability of existing statement positions as new statements were added to an instrument calibrated using the multi-facet model which measured people's impressions of a moisturising cream by touching its packaging. Twenty-one statements were developed using focus groups and online product reviews. Five products in tubes with different compliance characteristics were purchased and presented to 120 respondents, who were asked to squeeze, but could not see, the tubes. Respondents rated each tube on a five-point scale against 16 of the statements. The calibration of the statements showed that five of them indicated misfit to the model (i.e. they introduced high variance across the construct) and were removed from the analysis. After calibration, a different sample of 66 participants rated the tubes against the 11 calibrated statements to verify the consistency of scale and against 5 additional statements. During analysis of the data with the second set of statements added, two of the new statements were removed. The positions of the statement on the two scales were compared and all except one remained within measurement error of their original locations (Camargo and Henson 2013).

A stratified, four-fold, cross-validation study of responses to the second study demonstrated that statement locations were stable when analysed for different subsets of the sample. The whole sample of 186 participants was split into 4 approximately equal groups, and individuals were chosen at random from each group. The statements were recalibrated by selecting three groups forming a new sample for each n-calibration, leaving one group out, where n is the identification of the group combination used for calibration. The results of *t*-tests demonstrated a non-significant difference between the statement locations of the first 11 statements' set and each n-calibration. Furthermore, person locations from the first 11 statements' calibration and each n-calibration were highly correlated.

One of the benefits of using Rasch theory for measurement of affective responses is that it establishes a calibrated questionnaire, different administrations of which can be compared. The ability to compare the outcomes of studies has been demonstrated in the context of the evaluation of fabrics for vehicle seats (Henson and Camargo 2014), in which it was shown that affective responses to the fabrics were different depending on how the fabrics were presented to respondents. Product designers understand that product evaluation should take place in the setting and circumstances in which the product will be used. This concept is often referred to as the context by product designers, as ecological validity by psychologists (Brewer 2000) and as the frame of reference by Rasch researchers. However, there is often a trade-off between the internal and external validities of studies. The better the ecological validity of the experiment is, the harder it is to control variables, whereas the more the experiment is precisely controlled, the lower the external validity.

The study using fabrics for vehicle seats measured the effect of ecological validity by comparing affective responses to fabrics when touched as flat samples, and when touched as covers of vehicle seats. Three vehicle seat fabrics with different characteristics were used. In the first part of the study, 96 Japanese volunteers were asked to touch, but could not see, the flat pieces of fabric and rate them against statements (in Japanese) on a five-point Likert scale. In the second part of the study, the blindfolded participants were asked to touch the same fabrics used to cover three vehicle seats. In both parts of the study, a base of polyurethane foam with thickness of 3 mm was used under the fabrics. One scale was independently calibrated for each context. The locations of the fabrics on each scale were very similar, taking into account measurement errors. However, the locations of the statements on the two scales were uncorrelated. Similarly, paired *t*-tests showed that the person locations on the scale were significantly different in each context (Henson and Camargo 2014).

Using Rasch models to measure affective responses to products and then relate the measures to physical properties or features is novel. Klöcker et al. (2013) have applied Rasch theory to the measurement of the pleasantness of tactile textures. However, their rationale for using the unrestricted rating scale, when the analysis of responses of many people to many stimuli appears to demand a multi-facet model, is not clear. Their scale has been used to relate tactile pleasantness to measurements of friction (Klöcker et al. 2014).

Rasch measurement of affect has not yet been widely applied. Nevertheless, it has been demonstrated that Rasch measurement can be used to establish scales for measurement of affective responses. The novelty of the research in this paper is that it demonstrates that a scale can be used to relate affective responses to physical properties of a product, and consequently be used to design new products.

4. Method

4.1. *Introduction*

The aim of this experiment was to determine the correspondence between the physical properties of a product and affective responses, using Rasch theory to measure how the compliance of

Table 1. Everyday product packaging used to calibrate items of questionnaire.

Product	Packaging material
Baby food	130 ml laminated polypropylene, aluminium and polyethylene gusseted squeeze pouch with screw top closure
Toothpaste	150 ml laminated polypropylene, aluminium and polyethylene tube with flip-top closure
Hair Conditioner	200 ml low-density polyethylene tube with flip-top closure
Moisturiser	75 ml multi-layer low-density polyethylene and ethyl-vinyl-alcohol tube with flip-top closure
Baby bath lotion	300 ml oval, flat-based multi-layer low-density polyethylene and ethyl-vinyl-alcohol bottle with flip-top closure

packaging elicits an intuitive impression of a moisturiser cream it is said to contain. The hypothesis was tested that a new set of containers could be designed according to the compliances modelled on the basis of affective responses to the existing products. The forces participants applied when squeezing the packaging were measured to determine whether there is a correlation between the forces and the calibrated metric for affective responses.

The dimension intended to be measured was an affective response, denoted as 'impression' by the authors, of a face moisturising cream that the packaging was said to contain. In other words, the experiment tried to measure whether the packaging gave a favourable impression of the packaging's contents. Participants were not required to understand the concept of this construct or respond to the word 'impression'. This research was conducted with the approval of a Faculty Ethics Research Committee of the University of Leeds.

4.2. *Measuring compliance of product packaging*

The compliances of the five, commercially available, everyday products used in the previous study were measured (Camargo and Henson 2013) (Table 1). The compliance of each container was measured using a force table (MiniDyn, multi-component dynamometer Type 9256C2, Kistler) and an X–Z motion table (Series 1000 Cross Roller, Motion link). Each container was placed on the force table and pressed with a steel ball of radius 10 mm which was attached to the motion table. The measure of compliance was taken to be the displacement of the ball when the normal force was 3 N (Shao et al. 2010). Measurements were taken as the mean average of four repeated readings.

4.3. *Calibrated items*

The 14 statements calibrated in the previous study (Camargo and Henson 2013) were used (Table 2). The statements were developed from recorded verbatim statements by consumers during a focus group, along with other publicly available sources such as magazine advertisements. They were designed to capture aspects of the consumers' product experience. Table 3 shows the positions of the statements and packaging on the scale established in the previous study. The origin of zero was established as the default by the method used in the analysis. The default for the origin constrained the stimuli facet and the items facet at the centre of the logit scale with a mean of zero. Thus, solely the person facet floated on the continuum (not shown in Table 3). The locations of statements on the scale are indications of the ease of endorsement of the statements. The unit of the scales is the log-odds unit, the logit. One logit represents the distance along the scale that increases the odds of a respondent endorsing a statement by a factor of 2.718. Thus, the statements with the highest locations were less likely to be endorsed, and the products with the highest locations gave the best impression to participants of the moisturising cream they were said to contain.

Table 2. Set of calibrated statements used in this study.

ID	Statement	ID	Statement
1	The product in this container would give me a heavy, greasy film on my skin	8	The product in this packaging is likely to flow easily
2	The product in this container is likely to look and smell delightful	9	It is quite hard to explain the product when touching its packaging
3	I feel the product in this container would hydrate my skin	10	The product in this container could give me a refreshing sensation
4	The product in this packaging might be pricey	11	The product in this packaging could be a bit boring
5	The container makes me feel like I would be buying a great product	12	I could get just the right amount of the product when I squeeze its container
6	The product inside the container would spread easily	13	I feel this container as a skin care product
7	The product inside this container could be sticky	14	I could find no consistency in the product inside this container

Table 3. Mean locations of statements and packaging in the Rasch-calibrated log-linear scale.

Statement ID	Location (Logit)	SE (Logit)	Fit residual	Packaging ID	Location (Logit)	SE (Logit)	Fit residual	Compliance (mm/3 N)
6	– 0.88	0.20	0.19	4	1.23	0.23	0.18	4.11
2	– 0.55	0.22	0.26	3	0.59	0.21	0.14	4.74
12	– 0.46	0.20	0.26	2	– 0.17	0.20	0.46	5.70
14	– 0.38	0.18	0.58	1	– 0.25	0.19	0.29	6.08
10	– 0.18	0.21	0.16	5	– 1.40	0.18	0.46	1.02
1	– 0.17	0.21	0.38					
5	– 0.10	0.21	0.25					
13	0.02	0.18	0.33					
8	0.03	0.19	0.22					
3	0.31	0.20	0.25					
11	0.41	0.23	0.15					
7	0.44	0.19	0.33					
9	0.54	0.18	0.63					
4	0.96	0.21	0.32					

4.4. *Initial analysis of compliance measurements for packaging*

Based on the measures in Table 3, the degree of affective response does not follow the order of the magnitude of the compliance measurements for the packaging. A non-linear relationship between compliance values of the packaging and people's affective responses is observed (Figure 1). The regression indicated by the bold dashed line suggests three clusters of compliance values. The lower boundary of favourable impressions is located on the top of the graph, clustering Containers 3 and 4 within measurement error (standard error = 0.21 logit). The intersection of the regression line with the lower bound of relatively favourable responses determines the boundaries of lower and higher compliance between approximately 2.4 mm/3 N and 5.3 mm/3 N, indicating higher probability of endorsement that the containers provide a favourable impression of a moisturiser cream.

4.5. *Designing packaging for specific affective responses*

Using the range displayed in Figure 1, five packaging prototypes were designed with different values of compliance (Table 4). All prototypes were manufactured with the same dimensions, adopting a cylinder shape with the body diameter of 35 mm and height of 160 mm. A cap

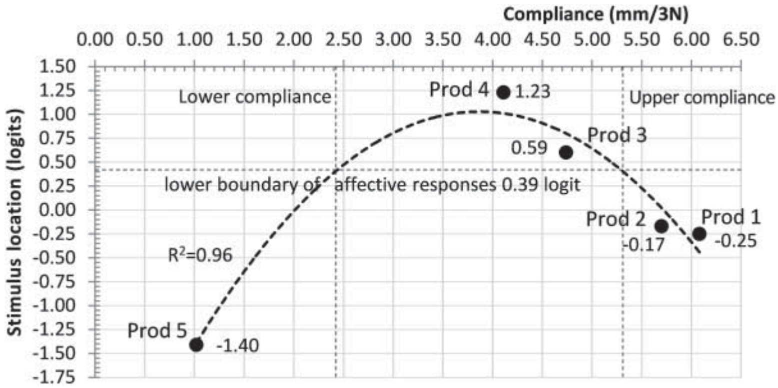
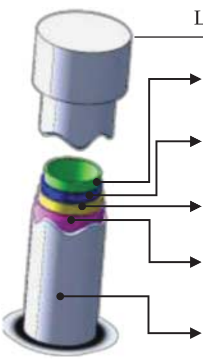


Figure 1. Non-linear regression determining a range of compliance values for favourable responses.

Table 4. Layers and materials used for the manufacture of prototypes.



Layer	R_a (μm)	Material	Type	Thickness (mm)
e	–	High-density polyethylene	Laminated sheet	0.743
d	–	Low-density polyethylene	Laminated sheet	0.406
c	–	Low-density polyethylene	Laminated sheet	0.306
b	–	Aluminium	Laminated foil	0.035
a	1.17	Polypropylene	Laminated film	0.060

with diameter of 46 mm and height of 52 mm was used to seal each container. All prototypes contained 139.5 cm^3 of the same moisturiser, filling about 90% of each container's internal volume.

The surface roughness of each tube was designed to be the same. The containers' surface roughness was measured using a stylus surface profilometer RTH Form Talysurf 120 L. The diamond stylus with radius 2.5 μm of the Talysurf machine scanned an area of 5 mm × 5 mm on the surface and recorded the peaks at a resolution of 1024 data points per mm^2. These were then filtered by the acquisition software to remove any apparent form. A texture profile in three dimensions was subsequently produced and post-processing software was used to determine the value of the arithmetical mean roughness, R_a (μm).

Layers of different materials were used to control the compliance of each prototype (Table 5). An acrylic adhesive was applied to the surface of the materials to bond the layers. The measurement of the layers' thickness was made with a micrometre screw gauge (Mitutoyo (0.001 mm)).

As in the previous study, the measure of compliance of the tubes was taken as the relative displacement of a 10-mm steel ball pressed into the packaging with a normal force of 3 N (Shao et al. 2010). The value taken was a mean of four readings.

The aim of manufacturing the new packaging was to determine whether the correspondence between compliance and affective response could be used to choose the compliance of new products. Thus, the scope for designing the multi-layer prototypes did not take into account

Table 5. Composition of the prototype packaging.

Stimulus	Prototype 1	Prototype 2	Prototype 3	Prototype 4	Prototype 5
Composition	a, b, c	a, b, c, c	a, b, d	a, b, c, e	a, b, d, e
Compliance mm/3 N	4.17	4.12	3.81	2.24	1.14

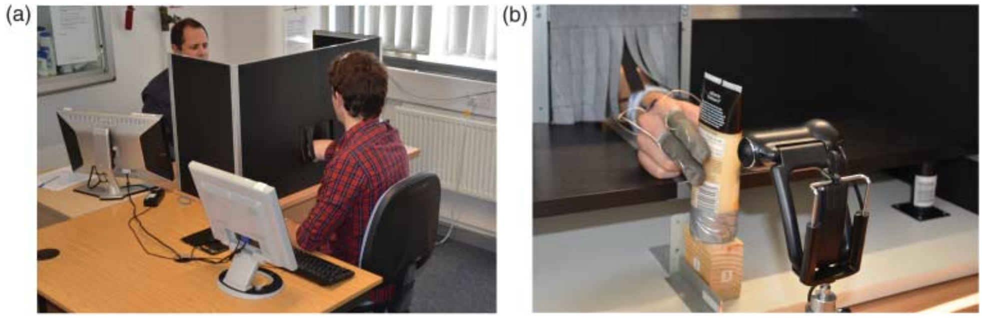


Figure 2. (a) Participant placing arm through a screen to touch the stimulus packaging and (b) arrangement for participants to squeeze stimulus packaging.

the nozzle arrangement, any barrier against oxygen ingress and aspects of sterilisation or aseptic filling. The layers were exclusively used to control the prototypes' compliance, and consequently the affective response they elicited from participants.

4.6. *Participants*

For this experiment, the statements had been previously calibrated, and consequently a sample of at least 64 participants was considered sufficient to produce the required standard error of less than 0.5 logit interval in a two-tailed 95% confidence interval of ± 2.0 (Linacre 1994).

The five prototypes were presented to 67 participants, 41.8% females and 58.2% males, 13.4% aged 18–25, 61.2% aged 26–35 and 25.4% over 35. Participants took part in the study one at a time. They were instructed to push their arm through a hole in a panel to touch the prototypes, one at a time, which were mounted vertically on wooden blocks (Figure 2(a)). Participants were instructed to touch the prototypes in the way that was most natural to them when squeezing ordinary containers. After squeezing each prototype participants rated their endorsement on a five-point Likert-style scale to the statements obtained in the previous calibration (Table 2) on computer-based self-report questionnaires. Participants were neither able to see the containers nor able to make contact with the moisturising cream inside them. The sequence in which the prototypes were presented was counterbalanced, and the sequence of items was randomised for each participant.

After squeezing the prototypes and responding to the questionnaires, participants were required to squeeze the containers once again, under the same conditions, while wearing tactile sensors (FingerTPS™ Wireless Tactile Force Measurement System) on their fingers to measure the forces applied (Figure 2(b)). The system was calibrated for each participant using a reference sensor at a force of 13.35 N. Participants squeezed each prototype twice with an interval of around two seconds between each touch, and the peak value was taken as the measure of interest. The sequence in which the prototypes were presented was counterbalanced. After completing a first sequence of prototypes, a second sequence was presented to the participants. Thus, four measurements were obtained from each participant for each prototype. The median value of the four peak values was calculated.

Results were standardised using Fisher's z-transformation and assessed for whether some observations exerted excessive influence on the tendency of the data. The effect of a single case on the whole regression model was indicated by Cook's distance (D_i), taking a value of $D_i \geq 0.06$ as a potential influential case (Cook 1979; Cook and Weisberg 1982). The difference between the predicted value when using an observation and the predicted value without using that observation was indicated by the statistical index DFFit (Belsey, Kuh, and Welsh 1980), where $\text{DFFit}_i \geq 0.24$ indicated an influential case for a sample size of 67. Finally, the covariance ratio (CVR) indicated the influence of an observation on the confidence interval (Belsey, Kuh, and Welsh 1980). Influential cases were flagged when $\text{CVR}_i \leq 0.91$, for a sample size of 67. Influential cases were identified and removed from the regression model.

4.7. *Recalibration of the scale*

The scale of the impression of the prototypes for a moisturiser cream based on their compliance was re-calibrated using the multi-facet Rasch model. The analysis was carried out using the protocol for calibration (Table 6) embodied in the software package RUMM2030® (Professional edition, 2012) (Andrich, Sheridan, and Luo 2012). The aim of Rasch analysis was to examine to what extent anomalies in the data corrupted measurement conditions.

Data were tested against the two assumptions of the Rasch model, response independence and trait independence (unidimensionality). Response dependency between statements was tested by observing whether the correlations of the item residuals were greater than the absolute value of

Table 6. Protocol for Rasch analysis in the study using the software package RUMM2030®.

Control	Definition
Invariance across the construct	Identification of overall discrepancy between the expected value and the observations across all statements based on a chi-square test for goodness of fit. Non-significant variance across the construct was indicated by $p > 0.05$ (Tennant, McKenna, and Hagell 2004)
Response dependence	Test for identifying whether a response to a statement in the scale interfered with the response to another statement. Response dependence between statements was examined by observing positive residual correlations (Tennant and Conaghan 2007)
Construct dependence (multi-dimensionality)	Examination of how well the data fitted together and cooperated to define the attribute being measured. This implied that after the extraction of the main scale, there was no remnant in the residuals' pattern resultant from relationships between statements except random associations (Smith 2002)
Standard error	Quantity added to or subtracted from a measure that gave the least distance required to establish a difference as statistically meaningful (Linacre 2005)
Standard residual	Indication of the difference between an observation and its expected value divided by the square-root of its modelled variance
Individual item-fit and individual person-fit	Indication of the degree of convergence between the expected value and observations for each person-item, typically equal to or lower than ± 2.5 (99% confidence interval). Variance within items was also examined through chi-square statistics. Persons holding extreme scores were not considered for calibration of scale (Tennant and Conaghan 2007)
Differential item functioning (DIF)	Test for identifying whether groups presented a consistently greater degree of endorsement to a statement for a prototype than another group (male and female groups and age groups) Two-way analysis of variance was applied for detecting significant DIF on the measures, allowing each element of the structure to be adjusted for any bias (Andrich and Hagquist 2004)
Thresholds ordering	Test for participants' inconsistency in using the response options carried out through identification of the thresholds patterns (Andrich 1985, 2010)
Target	Examination of whether the spread of statement thresholds on the continuum matched with the spread of persons. The average of item mean locations was centred at zero logit. Nevertheless, given that the attribute is represented by an interval scale, it could have been possible to re-score to a different range (e.g. 0–100) (Fisher 1992)

0.3 (Tennant and Conaghan 2007). Unidimensionality was tested through the method proposed by Smith (2002) of taking the factor loadings on the first residual component through a PCA to identify the two most divergent subsets of statements and then examining via a paired *t*-test any difference in the estimates that have been generated. Values ≤ 0.05 were taken as indications of unidimensionality (Smith 2002).

5. Results

5.1. *Calibration of the scale for the set of moisturiser prototypes*

The affective responses to the prototypes were analysed using the multi-facet Rasch approach. Individual item-fit residuals indicated no critical values (there were no standard residuals $> \pm 2.5$). All 14 statements fitted the model. Invariance across the measurement structure was indicated by a non-significant variance with chi-square probability $p = 0.51$. Table 7 shows the summary of fit statistics. Item–person residual correlations were lower than 0.3 and were therefore taken as an indication of non-significant response dependence. During the calibration, five persons were removed from the analysis as a consequence of extreme scores and high fit residuals.

The unidimensionality of the scale was confirmed using a binomial test, which indicated that fewer than 5% of the observations fell outside the *t*-range of ± 1.96 (95% confidence interval) (Smith 2002; Horton and Tennant 2010). No significant differential item functioning for sex and age was found. The statement and prototype locations on the scale are shown in Table 8.

The map of relative locations of all facets on the same logit scale is shown in Figure 3. Persons located towards the top of the scale were more inclined to endorse statements, whereas those at the bottom were less inclined. Statements located at the top of the scale were less likely to be

Table 7. Summary statistics of scale for the set of moisturiser prototypes.

Scale	Number of statements	Overall item-fit residual		Overall person-fit residual		Chi-square	df	*p*	*n*
		Mean	SD	Mean	SD				
Preliminary	14	0.48	0.51	– 0.06	2.55	191.21	140	> 0.05	67
Calibrated	14	0.32	0.43	– 0.06	2.15	138.83	140	0.51	62

Table 8. Mean location of prototypes and statements on the linear scale.

Stimulus	Mean location	SE	Statement	Mean location	SE
Prototype 1	0.89	0.20	14	– 1.32	0.19
Prototype 3	0.44	0.20	1	– 0.56	0.20
Prototype 2	0.23	0.20	12	– 0.28	0.20
Prototype 5	– 0.60	0.18	10	– 0.15	0.20
Prototype 4	– 0.96	0.20	3	– 0.15	0.18
			8	– 0.09	0.20
			9	– 0.09	0.21
			2	– 0.07	0.18
			4	– 0.03	0.19
			11	0.20	0.16
			13	0.31	0.20
			6	0.34	0.21
			5	0.86	0.18
			7	1.02	0.21

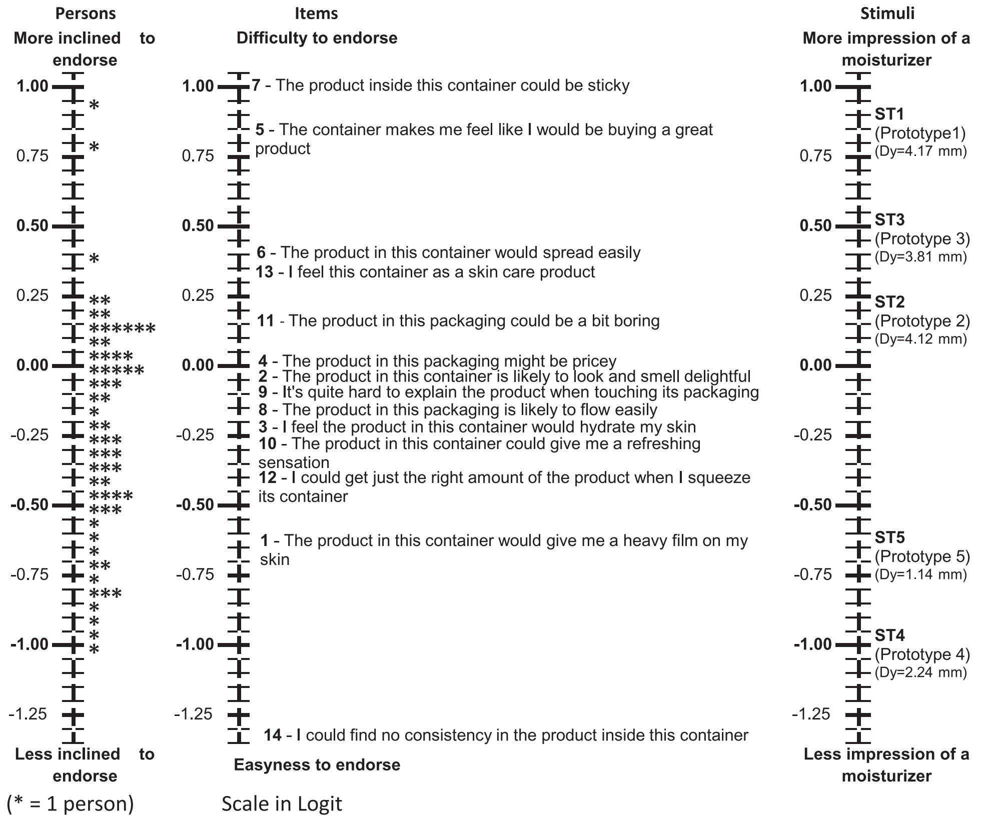


Figure 3. Rasch-calibrated metric with 14 statements for the impression of a moisturiser cream associated with the compliance of a range of product containers.

Table 9. Mean and standard error of the force applied on the prototypes without influential cases.

Stimulus	Prototype 1	Prototype 2	Prototype 3	Prototype 4	Prototype 5
Mean force (N)	22.3	22.6	23.5	26.1	37.4
Standard error (*N*)	2.8	3.1	3.3	3.3	6.3

endorsed. In the column Prototype, the prototypes at the top of the scale were more likely to be endorsed as packaging for a moisturiser than those at the bottom, according to participants' affective response.

5.2. *Sensory response when squeezing the container prototypes*

The z-transformation of the participants' applied forces demonstrated that some influential cases skewed the distribution of results, clustering at the lower forces and tailing towards the higher forces. Out of 335, 15 cases (5 stimuli × 67 participants) were indicated as influential following consideration in terms of Cook's distance, DFFit and CVR and were removed from the analysis. The resulting mean force and standard error for each prototype are shown in Table 9.

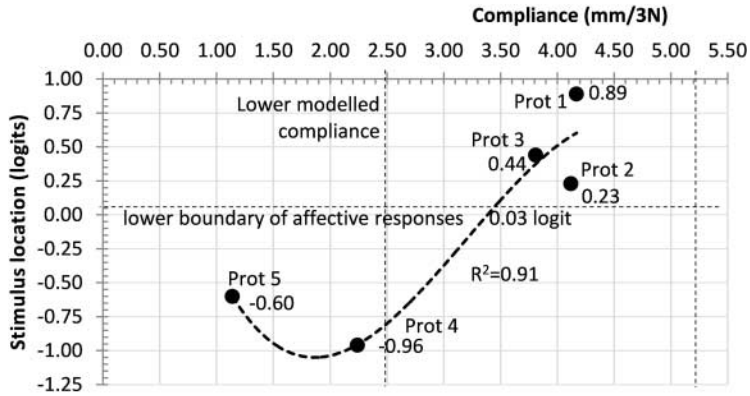


Figure 4. Correspondence between the compliance of the packaging and the intuitive impression of moisturiser cream the packaging is said to contain.

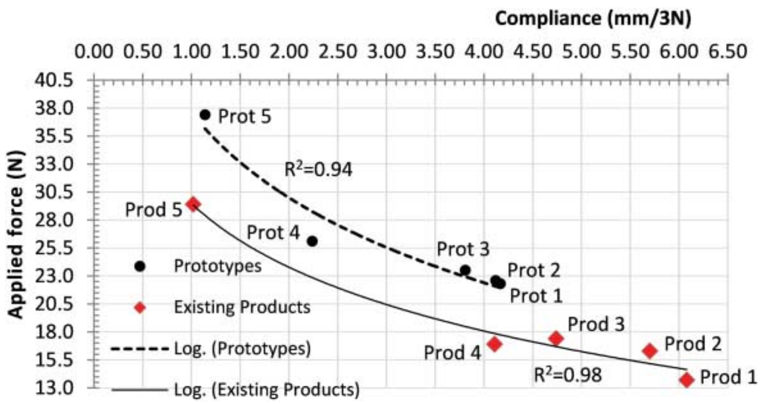


Figure 5. Comparison of the mean forces applied on the prototypes and existing products.

5.3. *Comparison between the prototypes and existing products*

The bold dashed line on the scatter plot in Figure 4 is the regression between the prototype locations on the log-linear continuum (Figure 3) and their measured compliance. It is observed that Prototypes 1, 2 and 3 clustered at the top of the graph with compliances between 3.81mm/3 N and 4.17 mm/3 N. These tubes are located within the lower and upper bounds of compliance for a favourable response as predicted by affective responses to existing products (Figure 1). Figure 5 also shows the compliances of 1.14 mm/3 N and 2.24 mm/3 N for Prototypes 5 and 4, respectively, fall below the lower bound, indicating less probability of a favourable response. It is noteworthy that the lower boundary for a more favourable affective response originates from the location of Prototype 2 plus its standard error (Table 8).

The regression line for the mean force applied on the moisturiser tubes presents a similar tendency as that applied on the existing products (Figure 5). Nevertheless, the range of force applied on the prototypes was higher than on the existing products. The regression line of Figure 5 indicates that the range of mean force applied on the existing products for a higher probability of a favourable affective response is between 16.9 and 17.4 N. For the prototypes the mean force lies between 22.3 and 23.5 N.

5.4. *Discussion of current study*

This study confirmed the prediction that in the context of packaging for moisturiser creams, container compliances of between 2.4 mm/3 N and 5.3 mm/3 N predicted in a previous study (Camargo and Henson 2013) would elicit the highest endorsement for giving a good impression of the product they were said to contain (Figure 4). Compliance was chosen as one of the physical measures of interest in this study because of its ease of measurement, its use in previous similar studies (Shao et al. 2010) and the reasonable assumption that there is a correspondence between compliance and people’s perception of softness or squeezability of the packaging. However, the perception of softness is more likely to be related to the way in which a deforming surface distributes pressure over the fingertip, such as the contact area spread rate (Bicchi, Scilingo, and de Rossi 2000). Compliance might not be the physical property with the best correlation with perceived softness or squeezability, and if a study such as this were repeated for the product development process of a commercial product, more variables likely to correspond to perceived softness would need to be included. Nevertheless, use of the measure of compliance appears to have produced good predictions of affective response.

The mean force applied on the existing products was consistently lower than the force applied on the prototypes for all levels of relative compliance (Figure 5). This could be because of the specific contact area spread rate of the prototypes. It might be a consequence of the influence of different factors other than the containers’ compliance when the persons processed the sensory information. The containers’ characteristics of shape, for example, might have been combined with the characteristic of compliance for the respondents to form clear mental representations based on the tactile sensory information (d’Astous and Kamau 2010). Therefore, the participants might have had fewer indications about the impression of a moisturiser cream when they squeezed the prototypes than the existing products although this can have varied from person to person. Other cues present in the real products which were removed in the prototypes, such as large variations in thermal properties, might have influenced affective responses (Chen et al. 2009). These hypotheses require further investigation.

Previous research demonstrated that Rasch measurement theory can be used to create stable, calibrated scales of affective response, based on many people’s ratings of many products. This research further extended knowledge of the novel application of Rasch theory in product design, by showing that the scales can be used to relate the physical properties of products with people’s affective responses. The scope of the study was relatively small, in that only compliance and applied touching force were chosen as the physical measures, whereas a typical kansei study might have attempted to correlate responses with many physical features and measures. However, it can be argued that experiments such as these should be based on hypotheses established before experimentation begins, rather than seeing what emerges from the data, so that independent variables are chosen that can be reasonably argued to influence affective response. This experiment might have been improved had it been possible to design packaging with predefined compliances, rather than crafting prototypes and then measuring their compliances. However, this does not invalidate this study’s findings.

While there appear to be some reasons to prefer analysing responses to self-report data using the Rasch approach where appropriate, rather than using PCA, this study is not a comparison of the two approaches. Each approach has its advantages in different circumstances. In particular, if an exploratory, multi-dimensional semantic space is needed, the PCA approach is more appropriate, whereas the Rasch approach is perhaps more appropriate for measurement along unidimensional scales. However, the authors are of the view that the use of adjective pairs should be avoided, because of the lack of a rationale that the collection of words represents an underlying measurement construct. This point is discussed further in Section 6.

6. General discussion of Rasch approach in product design

The criticism of the use of adjective pairs made in Section 5, that the assumption that responses to adjectives pairs constitute an underlying measurement structure is weak, can be extended to the Rasch approach. Challenges include whether responses that fit the model can be taken as evidence that an underlying construct exists, and how to interpret additivity in the context of a latent variable (Kline 1998; Sijtsma 2010). It can be argued that the verbatim statements used on the questionnaires, which are derived from consumer research, possess some face validity and that responses to them will reflect an underlying construct. This argument cannot be applied as easily to adjective pairs. Nevertheless, there is currently little reason for choosing one statement over another to include in a questionnaire, and while a rationale for explaining why a statement is rejected by the analysis can be hypothesised after the fact, it would be an advantage to be able to select ‘good’ statements before experimentation begins. For one thing, trialling statements that do not work is expensive, although the main point is to have a scientific rationale for developing and selecting statements. Research using computational and corpus linguistics might provide practical tools to address this problem.

Researchers who use physiological measures of affective response might argue that the problem can best be avoided by not using self-report questionnaires because, additionally, people cannot be trusted to respond honestly to questionnaires. The same problem exists here, however, because of a lack of a rationale for assuming that changes in physiological measures represents some underlying affective response that can be measured. There is no theoretical reason, though, why a Rasch-based instrument cannot include physiological and behaviour items instead of, or as well as, self-report items.

Related to the problem of whether responses to statements can reflect an underlying latent construct associated with affective responses to products is whether consumer preferences can even have an underlying quantitative structure (Salzberger 2013). One issue is to do with ranking. Shown a set of real products, such as prestige motor vehicles, different men from a well-targeted sample demographic or attitudinal group would be likely to rank their preferred vehicles differently (Köster 2003), in which case there cannot be a single scale for the whole sample. Research by the authors has not yet confronted this problem. One would have expected to observe different ranked preferences in their study of confectionery (Camargo and Henson 2011) had they been looked for. That the responses of around 10% of participants were removed for calibration of the scale because they did not fit the model did not seem unreasonable, and it does not seem to point to the issue of ranked preferences being a serious issue. It might be that the problem can be adequately addressed through differential item functioning, where different ranked preferences form groups for analysis. A further issue is whether the assumption of stimulus independence can hold when people are expressing evaluations of many products, because of the asymmetric domination effect, in which it has been shown that people’s preferences between two alternatives are affected by the relative attributes of a third option (Huber, Payne, and Puto 1982).

It is yet to be established whether there are any differences between the scales developed using the classical, statistical approach and those developed using Rasch measurement theory (Nunnally 1978). The linearity of the responses and the anticipated improvement in accuracy are likely to be beneficial in applications requiring quantification of latent traits, such as when relating emotional or affective responses to product experience to measures of the physical properties of products. However, these are not necessarily the principal reasons for choosing the Rasch approach in the context of product design. The benefits of the Rasch approach are the ease with which outcomes of studies can be compared, and the ability to develop instruments that can be administered reliably to small samples (Embretson and Reise 2000). These will enable, for example, the comparison of affective responses to products between samples of consumers from different cultural, demographic or attitudinal groups, and between consumers and expert panels. They should lead to the implementation of CATs of standardised questionnaires, which could substantially reduce the costs of measuring consumers' affective responses to product designs.

7. Conclusions

Rasch measurement theory offers an alternative approach for developing scales of people's affective responses to products. Its use is well established in other disciplines, but its application to product design is novel. The research reported here used Rasch measurement theory to measure the affective impression of a moisturiser cream elicited by the compliance of the moisturiser's packaging, and used the derived information to create new designs that elicited predictable affective responses from consumers. The benefits of using the Rasch approach are in the ability to obtain reliable results from small samples of consumers, once a set of statements has been calibrated, thus reducing the costs of affective design. While the research shows that the use of Rasch measurement theory is a viable approach, it remains to be demonstrated that it can deliver the anticipated business benefits.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. Advertising slogan used by the Gillette Safety Razor Company.
2. Phrase used in relation to laundry detergent by both P&G and Colgate.

References

Andrich, D. 1978. "Application of a Psychometric Rating Model to Ordered Categories Which are Scored with Successive Integers." *Applied Psychological Measurement* 2 (4): 581–594.
Andrich, D. 1985. "An Elaboration of Guttman Scaling with Rasch Models for Measurement." In *Sociological Methodology*, edited by N. Brandon-Tuma, 33–80. San Francisco, CA: Jossey-Bass.
Andrich, D. 1988. *Rasch Models for Measurement, Sage University Papers Series on Quantitative Applications in the Social Sciences*. 68 vols. London: Sage.
Andrich, D. 2010. "Sufficiency and Conditional Estimation of Person Parameters in the Polytomous Rasch Model." *Psychometrika* 75 (2): 292–308.
Andrich, D., and C. Hagquist. 2004. "Detection of Differential Item Functioning using Analysis of Variance." 2nd International conference on Measurement in Health, Education, Psychology and marketing: Developments with the Rasch Models, Perth, Australia, January 20–22.
Andrich, D., B. E. Sheridan, and G. Luo. 2012. "RUMM2030: Rasch Unidimensional Models for Measurement." Perth: RUMM laboratory.
Belsey, D. A., E. Kuh, and R. Welsh. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
Bicchi, A., E. P. Scilingo, and D. de Rossi. 2000. "Haptic Discrimination of Softness in Teleoperation: The Role of the Contact Area Spread Rate." *Robotics & Automation, IEEE Transactions* 16: 496–504.
Borsboom, D. 2006. "When Does Measurement Invariance Matter?" *Medical Care* 44 (11): 176–181.
Brewer, M. 2000. "Research Design and Issues of Validity." In *Handbook of Research Methods in Social and Personality Psychology*, edited by H. Reis and C. Judd, Chap. 1, 3–16. Cambridge: Cambridge University Press.
Camargo, F. R. 2013. "Calibrating Scales for Affective Responses to Design Elements using Rasch Measurement Theory." PhD diss., School of Mechanical Engineering, University of Leeds, UK.
Camargo, F. R., and B. Henson. 2011. "Measuring Affective Responses for Human-Oriented Product Design Using the Rasch Model." *Journal of Design Research* 9 (4): 360–375.
Camargo, F. R., and B. Henson. 2012. "The Rasch Probabilistic Model for Measuring Affective Responses to Product Features." *International Journal of Human Factors and Ergonomics* 1 (2): 204–219.

Camargo, F. R., and B. Henson. 2013. "Aligning Physical Elements with Persons' Attitude: An Approach Using Rasch Measurement Theory." *Journal of Physics: Conference Series* 459. doi:10.1088/1742-6596/459/1/012009
Chen, X., F. Shao, C. J. Barnes, T. Childs, and B. Henson. 2009. "Exploring Relationships between Touch Perception and Surface Physical Properties." *International Journal of Design* 3 (2): 67–77.
Childs, T., V. Agouridas, C. Barnes, and B. Henson. 2006. "Controlled Appeal Product Design: A Life Cycle Role for Affective (Kansei) Engineering." Proceedings of the 13th CIRP International Conference on Life Cycle Engineering, Leuven, Belgium, May 31–June 2.
Cook, D. 1979. "Influential Observations in Linear Regression." *Journal of the American Statistical Association* 74 (36): 169–174.
Cook, D., and S. Weisberg. 1982. *Residuals and Influence in Regression*. New York: Chapman & Hall.
d'Astous, A., and E. Kamau. 2010. "Consumer Product Evaluation Based on Tactile Sensory Information." *Journal of Consumer Behaviour* 9 (3): 206–213.
Drake, M. A., D. Drake, F. Bodyfelt, S. Clark, and M. Costello. 2009. "History of Sensory Analysis." In *The Sensory Evaluation of Dairy Products*, edited by S. Clark, M. Costello, M. Drake, and F. Bodyfelt, 1–6. New York: Spring-Verlag.
Embretson, S., and S. Reise. 2000. *Psychometric Methods: Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
Fisher, W. P. 1992. "Reliability Statistics." *Rasch Measurement Transactions* 6 (3): 238.
Gunes, H., M. Piccardi, and M. Pantic. 2008. "From the Lab to the Real World: Affect Recognition Using Multiple Cues and Modalities." In *Affective Computing, Focus on Emotion Expression, Synthesis and Recognition*, edited by J. Or, 185–218. Vienna: I-Tech Education and Publishing.
Heise, D. R. 1969. "Some Methodological Issues in Semantic Differential Research." *Psychological Bulletin*, 72 (6): 406–422.
Henson, B., and F. R. Camargo. 2014. "Verification of Quantitative Properties of Kansei Scales Using the Rasch Model." *Journal of the Japan Society of Kansei Engineering* 13 (1): 13–20.
Horton, M., and A. Tennant. 2010. "Assessing Unidimensionality using Smith's (2002) Approach in RUMM 2030." Proceedings of the International Conference on Probabilistic Models for Measurement in Education, Psychology, Social Science and Health, Copenhagen, June 13–16.
Hotta, H., and M. Hagiwara. 2005. "An Automatic Rule Creating Method for Kansei Data and Its Application to a Font Creating System." Proceedings of the 2nd Conference on Modelling Decisions for Artificial Intelligence, Tsukuba, Japan, July 25–27.
Huber, J., J. W. Payne, and C. Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9 (1): 90–98.
Ishihara, S., K. Ishihara, M. Nagamachi, and Y. Matsubara. 1995. "An Automatic Builder for an Engineering Expert System Using Self-Organizing Neural Networks." *International Journal of Industrial Ergonomics* 15 (1): 13–24.
Karabatos, G. 2001. "The Rasch Model, Additive Conjoint Measurement, and New Models of Probabilistic Measurement Theory." *Journal of Applied Measurement* 2 (4): 389–423.
Kingdom, F., and N. Prins. 2009. *Psychophysics: A Practical Introduction*. London: International Academic Press. ISBN: 978-0-12-373656-7.
Kline, P. 1998. *The New Psychometrics: Science, Psychology and Measurement*. London: Routledge.
Klöcker, A., C. Arnould, M. Penta, and J-L. Thonnard. 2013. "Rasch-Built Measure of Pleasant Touch through Active Fingertip Explorations." *Frontiers in Neurorobotics* 6 (5): 1–9.
Klöcker, A., C. M. Oddo, D. Camboni, M. Penta, and J-L. Thonnard. 2014. "Physical Factors Influencing Pleasant Touch During Passive Fingertip Stimulation". *PLoS ONE* 9 (7): e101361. doi:10.1371/journal.pone.0101361
Köster, E. P. 2003. "The Psychology of Food Choice: Some Often Encountered Fallacies." *Food Quality and Preference* 14 (5–6): 359–373.
Likert, R. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 140: 1–55.
Linacre, J. M. 1989. *Many-Facet Rasch Measurement*. Chicago: Mesa Press.
Linacre, J. M. 1994. "Sample Size and Item Calibration Stability." *Rasch Measurement Transactions* 7 (4): 328.
Linacre, J. M. 2002. "Facets, Factors, Elements and Levels." *Rasch Measurement Transactions* 16 (2): 880.
Linacre, J. M. 2005. "Standard Errors: Means, Measures, Origins and Anchor Values." *Rasch Measurement Transactions* 19 (3): 1030.
Luce, R. D., and J. W. Tukey. 1964. "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement." *Journal of Mathematical Psychology* 1 (1): 1–27.
Masters, G. N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (2): 149–174.
Mordkoff, A. M. 1963. "An Empirical Test of the Functional Antonymy of Semantic Differential Scales." *Journal of Verbal Learning and Verbal Behavior* 2 (5–6): 504–508.
Nagamachi, M. 1995. "Kansei Engineering: A new Ergonomic Consumer-Oriented Technology for Product Development." *International Journal of Industrial Ergonomics* 15 (1): 3–11.
Nagamachi, M. 2008. "Perspectives and the New Trend of Kansei/Affective Engineering." *The TQM Journal* 20 (4): 290–298.
Nunnally, J. O. 1978. *Psychometric Theory*. New York: McGraw-Hill.
Osgood, C., G. Suci, and P. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Rasch, G. 1960 [1980]. *Probabilistic Models for Some Intelligence and Attainment Tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by Wright, B. D. Chicago: The University of Chicago Press.

Rasch, G. 1961. "On General Laws and the Meaning of Measurement in Psychology." Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 321–334, June 20–July 30.

Rossi, G. B. 2007. "Measurability." *Measurement* 40 (6): 545–562.

Salzberger, T. 2009. *Measurement in Marketing Research: An Alternative Framework*. Cheltenham: Edward Elgar International. ISBN 9781848441651.

Salzberger, T. 2013. "Attempting Measurement of Psychological Attributes." *Frontiers in Psychology* 4 (75): 1–4.

Schaaff, K. 2009. "Towards Emotion Recognition from Electroencephalographic Signals." Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, Amsterdam, September 10–12.

Shao, F., X. Chen, C. J. Barnes, and B. Henson. 2010. "A Novel Tactile Sensation Measurement System for Qualifying Touch Perception." *Proceedings of the Institution of Mechanical Engineers, Part H, Journal of Engineering in Medicine* 224 (1): 97–105.

Shiffman, S. S., M. L. Reynolds, and F. W. Young. 1981. *Handbook of Multidimensional Scaling*. New York: Academic Press.

Sijtsma, K. 2010. "Psychological Measurement between Physics and Statistics (Keynote)." Proceedings of the 4th International Conference on Probabilistic Models for Measurement in Education, Psychology, Social Science and Health, Copenhagen, June 13–16.

Smith, E. V. 2002. "Detecting and Evaluating the Impact of Multidimensionality Using Item fit Statistics and Principal Component Analysis of Residuals." *Journal of Applied Measurement* 3 (2): 205–231.

Snider, J. G., and C. E. Osgood. 1969. *Semantic Differential Technique: A Sourcebook*. Chicago: Aldine.

Stevens, S. S. 1946. "On the Theory of Scales of Measurement." *Science* 103: 677–680.

Tennant, A., and P. G. Conaghan. 2007. "The Rasch Measurement Model in Rheumathology: What is it and Why use it? When Should it be Applied, and What Should One Look for in a Rasch Paper?" *Arthritis & Rheumatism* 57 (8): 1358–1362.

Tennant, A., S. P. McKenna, and P. Hagell. 2004. "Application of Rasch Analysis in the Development and Application of Quality of Life Instruments." *Value in Health* 7 (1): S22–S26.

Thomeé, R., G. Grimby, B. D. Wright, and J. M. Linacre. 1995. "Rasch Analysis of Visual Analog Scale Measurements Before and After Treatment of Patellofemoral Pain Syndrome in Women." *Scandinavian Journal of Rehabilitation Medicine* 27 (3): 145–151.

Wright, B. D., and S. R. Bell. 1984. "Item Banks: What, Why, How." *Journal of Educational Measurement* 21 (4): 331–345.

Wright, B. D., and J. M. Linacre. 1989. "Observations are Always Ordinal; Measurements, However, Must be Interval." *Archives of Physical Medicine and Rehabilitation* 70 (12): 857–860.

Zeng, Z., M. Pantic, G. I. Roisman, and T. S. Huang. 2009. "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (1): 39–58.