Aligning physical elements with persons' attitude: an approach using Rasch measurement theory

2013 J. Phys.: Conf. Ser. 459 012009

(http://iopscience.iop.org/1742-6596/459/1/012009)

Aligning physical elements with persons' attitude: an approach using Rasch measurement theory

F R Camargo[1] and B Henson

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

[1]E-mail: mnfrc@leeds.ac.uk

Abstract. Affective engineering uses mathematical models to convert the information obtained from persons' attitude to physical elements into an ergonomic design. However, applications in the domain have not in many cases met measurement assumptions. This paper proposes a novel approach based on Rasch measurement theory to overcome the problem. The research demonstrates that if data fit the model, further variables can be added to a scale. An empirical study was designed to determine the range of compliance where consumers could obtain an impression of a moisturizer cream when touching some product containers. Persons, variables and stimulus objects were parameterised independently on a linear continuum. The results showed that a calibrated scale preserves comparability although incorporating further variables.

1. Introduction

Research in affective engineering has focused on converting the information obtained from people's feelings (e.g., emotions, feelings, sentiments and moods) stimulated by the physical properties of materials or product features (e.g., roughness, temperature, odour and colour) into an improved design. However, current methods in the domain have not in many cases met measurement assumptions. The relation amongst variables in a measurement structure ought, in this sense, to present additivity, invariant comparisons and a constant unit [1]. Some sources of violation to those assumptions have been associated with linguistic ambiguity, redundancy, misinterpretation, variance emerged from bias, reliability of qualitative dimensions and the difficulties of assuming equal intervals in categorical scales [2]. Additionally, some approaches have not embodied metrological rules, such as those of traceability and variance control [3]. As a consequence, one cannot infer beyond the considered sample and results from different studies cannot reliably be compared, limiting the understanding of a more general human–product interaction.

This paper proposes overcoming the problem by establishing a novel theory-driven approach in the domain of affective engineering using probabilistic models underpinned by Rasch measurement theory (RMT). The Rasch model (RM) provides theory and mechanisms to examine how well the data fit together and cooperate to define the attribute being measured. Accordingly, the RM's procedures verify whether the observations meet the assumptions necessary for quantifying the numerical validity of the data employing the tools of standard statistics.

The research demonstrates through an empirical study that if data from affective responses to physical elements fit the model, further variables can be added to the calibrated core of measurement whilst preserving comparability. The study established a scale to compare affective responses indicating different degrees of endorsement associated with the tactile sensory information when squeezing a collection of product containers with different characteristics of compliance. The aim was to determine the range of compliance where consumers could obtain an impression of a moisturizer cream. Persons, variables (i.e., statements in a questionnaire, called items henceforth) and stimulus objects were independently parameterised on a sole linear continuum. This alignment allowed the inclusion of further items in the scale, re-calibrating it through the responses from a different group of individuals.

2. Separation of parameter estimates

As a result of the additive correspondence obtained from a calibrated metric, comparisons can be made by the difference between the numbers associated with the responses, where a particular difference has the same interpretation across the scale continuum.

From RMT's perspective, separation of parameters estimates is a determinant condition for achieving invariant comparisons. The Rasch model, named after the Danish mathematician George Rasch, expresses the probability that a person will endorse an item with two-category responses (e.g., yes or no) as a logistic function of the difference between two mathematically independent parameters on a linear continuum: the persons' parameter and the items' parameter [4]

$$\Pr\{X_{ni} = 1|\beta,\delta\} = \frac{e^{(\beta_n-\delta_i)}}{1+e^{(\beta_n-\delta_i)}} \tag{1}$$

where $\Pr\{X_{ni} = 1|\beta,\delta\}$ is the probability that a person *n* will endorse an item *i*, β is the persons' parameter and δ is the items' parameter. A number of models have extended the dichotomous Rasch model for accommodating items with more than two categories [5][6]. We have adapted a derivation of the Rasch model [7], namely the many-facet Rasch model (MFRM) developed by Linacre [8], which embodies one or more parameters called facets to the RM' structure. In the domain of affective engineering we have used the facet denoted stimulus, associating it with physical elements of design, such that

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

given that $\psi_{nis} = \sum_{k=0}^{m}\exp\left[\sum_{k=0}^{x}(\beta_n - \delta_i + \zeta_s - \tau_{isk})\right]$ is a normalising factor and $\Pr\{X_{nis} = x_{nis}|\beta,\delta,\zeta\}$ is the probability of a respondent *n* to give a rating of k, $k \in (1,\ldots,m)$, on item *i* for stimulus *s;* β_n is the inclination of a person *n* to endorse the item *i* for stimulus *s;* ζ_s is the level of the attribute fulfilment for stimulus *s;* δ_i is the difficulty of endorsement of item *i* and τ_{isk} is the threshold parameter given a rate *k* on item *i* for stimulus *s*.

3. Empirical study - Method

3.1. Preliminary pool of statements and sampling

Initially, documented verbatim statements to express affective requirements were collected from a focus group along with other publicly available sources. Sixteen preliminary statements related to the packaging of everyday products were identified and selected according to the context of the study (Table 1). Subsequently, five available products in the market with different characteristics associated with the compliance of their containers were presented to 120 respondents. They were neither able to see the containers nor able to make contact with the product inside them. After squeezing each container participants expressed their degree of endorsement on a five-point Likert-style scale using computer-base self-report questionnaires. The metric was then calibrated. Finally, five additional items were introduced in the calibrated metric (Table 2). A different sample of 66 participants gave their ratings on the set of statements with the additional items. The same containers presented to the previous sample were used in this administration. A new calibration was then carried out and compared with the first test.

Table 1. Preliminary pool of statements for the first administration

Code	Item	Code	Item
1	The product in this container would give me a heavy, greasy film on my skin.	9	There is a lightweight cream in this container.
2	The product in this container is likely to look and smell delightful.	10	It is easy to know how much is left in the packaging.
3	I might get a bit watery product in this container.	11	The product inside this container could be sticky.
4	I feel the product in this container would hydrate my skin.	12	The product in this packaging is likely to flow easily.
5	The product in this packaging might be pricey.	13	The product in this packaging might seem more medicinal than anything else.
6	The container feels only half filled when squeezing it.	14	It is quite hard to explain the product when touching its packaging.
7	The container makes me feel like I would be buying a great product.	15	The product in this container could give me a refreshing sensation.
8	The product inside the container would spread easily.	16	The product in this packaging could be a bit boring.

Table 2. Additional statements for the second administration

Code	Item	Code	Item
17	I could get just the right amount of the product when I squeeze its container.	20	It’s too soft for a creamy product.
18	I’ve got a pleasant touch with this container.	21	I could find no consistency in the product inside this container.
19	I feel this container as a skin care product.		

3.2. *Rasch analysis*

The investigation using Rasch analysis determined whether the data fitted the model. The analysis’ procedures, therefore, contrasted with the practice in statistical modelling, which attempts to fit a model to data. The residuals were examined to determine the difference between the observed score and the expected score for each person for each item. Adequate fit to the model was established within the theoretical interval of ±2.50, which represents approximately 99% of the confidence interval (CI) [9].

Items were tested for misfit associated with the respondents’ inconsistent use of the response categories. This involved examining whether the transitions between categories, called thresholds, were consistent, i.e., whether each response category had a point along the scale continuum where the most probable response was located. Another procedure was to examine item bias for sex and age, denoted as differential item functioning (DIF), i.e., whether a person subgroup demonstrated consistently greater inclination to endorse an item than another subgroup. Violation of the assumption of local independency, which underpins the RM, was examined through of the degree of correlation in which the response to an item determined the response to another one. In addition, test for trait independence, also called unidimensionality, determined whether after the calibration there was no significant pattern in the residuals resultant from relationships between items except random associations. The person separation index (PSI) is another indication of quality control of the calibration associated with the reliability of the metric. However, in RMT the precision of the individual estimates is emphasised and therefore, the index was solely used as an element of a comprehensive interpretation of the data set.

3.3. *Measurement of the containers’ compliance*

The testing system to measure compliance consisted of a force platform (MiniDyn, multi-component dynamometer Type 9256C2, Kistler), an X–Z motion table (Series 1000 Cross Roller, Motion link), a steel ball of radius 10mm, a controller and a computer program. The containers were positioned between the steel ball and the force platform. The ball was pressed against the surface of each stimulus and the ball’s displacement D_y with increasing load F_y was recorded. The measure of compliance was empirically taken to be the value of D_y (mm) when F_y was 3N [10].

4. Results

4.1. Calibration of the metric using the first sample

A preliminary analysis using the software package RUMM2030®, professional edition [11], indicated high residual correlations and evidence of multidimensionality in the structure. Disordered thresholds were identified in every item. The score system was then recoded into four categories and Items 1, 11, 14 and 16 were given reversed scores. Examination of the correlation matrix of person-item residuals identified items with indication of dependency. Misfit to the model was identified for Items 3, 6, 9, 10 and 13, being removed from the analysis. No DIF was observed. The remaining items fitted the model, indicating relative independence and meeting the model's assumption of unidimensionality (Table 3).

The metric presented in Figure 1 is the representation of the relative locations of all facets on the same logit scale. Person locations are plotted on the scale represented in the first column. On the top of the second column are indicated items that obtained endorsement by fewer persons. The location of stimuli on the continuum demonstrated that the container with lower compliance was located at the bottom of the scale in relation to the containers with higher compliance, indicating higher degree of endorsement to the affective attribute for the latter.

Table 3 - Fit statistics of the scale for the first sample.

Stimulus	Location	*SE*	Fit-residual	Items	Mean location	*SE*	Fit-residual	χ^2	df	*p*	PSI
St1	0.03	0.14	0.04	1	-0.04	0.14	0.37	121.64	110	0.21	0.7
St2	-0.07	0.13	0.07	2	-0.25	0.15	-0.08				
St3	0.34	0.15	-0.04	4	0.00	0.13	-0.06				
St4	0.28	0.15	0.09	5	-0.22	0.15	-0.03				
St5	-0.58	0.14	0.27	7	0.04	0.15	0.02				
				8	-0.50	0.13	0.04				
				11	0.43	0.14	0.31				
				12	-0.31	0.14	0.12				
				14	0.27	0.13	0.44				
				15	-0.02	0.14	-0.05				
				16	0.60	0.16	-0.06				

4.2. Calibration of the metric with additional items using the second sample

The calibrated items from the first administration were used as the core of a second calibration with the five incorporated items. In a preliminary analysis significant item-trait interaction and high residual correlations were identified, giving evidence of misfit to the model. Items were then rescored into four categories and reversed for Items 20 and 21. Items 18 and 20 were removed from the preliminary set. After calibration, a non-significant item-trait interaction (p=0.96) indicated invariance across the structure (Table 4). Figure 2 is the representation of the relative locations of the calibrated 14-item set.

Table 4 - Fit statistics of the scale for the second sample.

Stimulus	Location	*SE*	Fit-residual	Items	Mean location	*SE*	Fit-residual	χ^2	df	*p*	PSI
St1	-0.25	0.19	0.29	1	-0.17	0.21	0.38	112.27	140	0.96	0.8
St2	-0.17	0.20	0.46	2	-0.55	0.22	0.26				
St3	0.59	0.21	0.14	4	0.31	0.20	0.25				
St4	1.23	0.23	0.18	5	0.96	0.21	0.32				
St5	-1.40	0.18	0.46	7	-0.10	0.21	0.25				
				8	-0.88	0.20	0.19				
				11	0.44	0.19	0.33				
				12	0.03	0.19	0.22				
				14	0.54	0.18	0.63				
				15	-0.18	0.21	0.16				
				16	0.41	0.23	0.15				
				17	-0.46	0.20	0.26				
				19	0.02	0.18	0.33				
				21	-0.38	0.18	0.58				

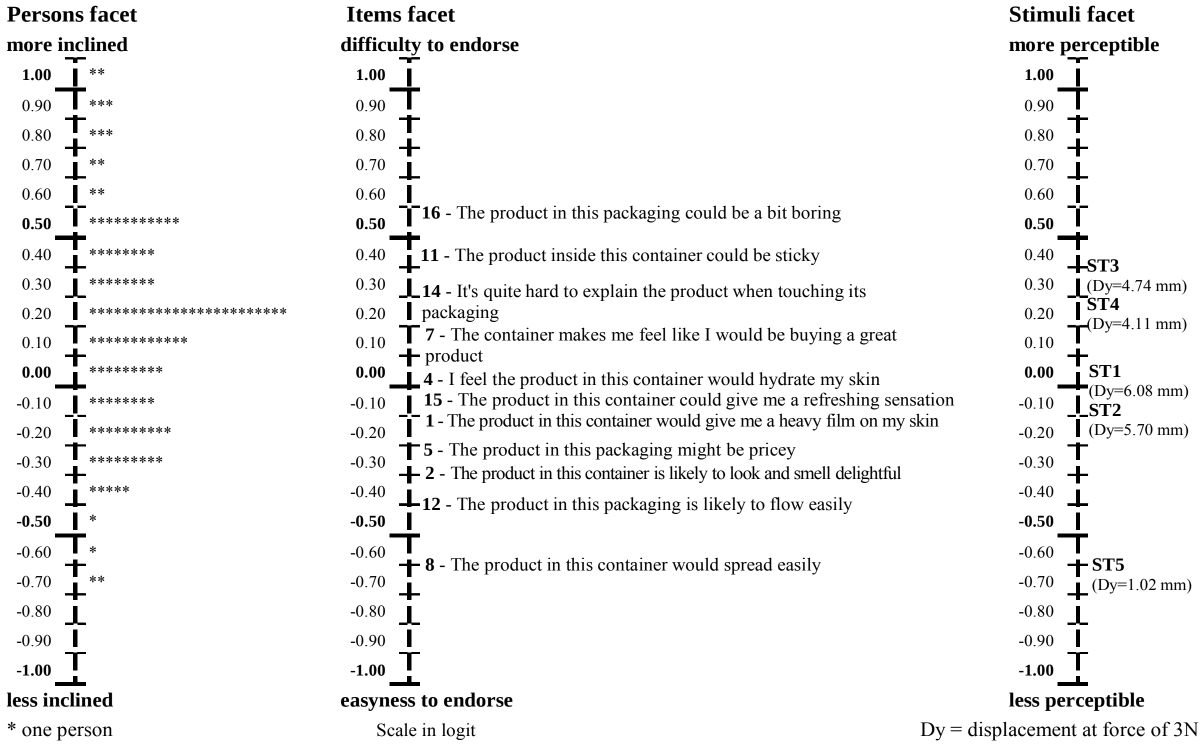

Figure 1. Rasch-calibrated metric with 11 items for the impression of a moisturizer cream associated with the compliance of a range of product containers.

5. Discussion

In Rasch analysis items that do not fit the model are not automatically rejected; rather they are investigated to identify their source of misfit and to what extent they corrupt measurement. For example, responses to Item 3, “I might get a bit watery product in this container”, presented dependence on Item 12 and Item 18, “I’ve got a pleasant touch with this container”, can have been ambiguous originating an inconsistent responses pattern.

The inclusion of further items in the scale yielded a better discrimination of the effects of the compliance of the containers for that particular sample. This was not unexpected because more information was aggregated to the measurement structure.

Although item locations on the continuum did not achieve identical estimates, these differences can be associated with the measurement error in the compared calibrations. The modelled standard error was estimated to be lower than 0.5 logit interval in a two-tailed 95% confidence interval of ± 2.0, based on the calibration stability for the sample size adopted in the study. An exception was Item 5 that extrapolated its measurement error and therefore, it shall be further examined. Nevertheless, the measurement precision of the calibration will normally improve as more data become available.

The metrics have suggested that the degree of affective responses does not follow the order of the physical element compliance for the containers. The metrics indicate, however, that according to participants’ impression there is a higher likelihood of the intermediate range of compliance being associated with a container of a moisturizer cream. It is possible to draw this conclusion because of the alignment of persons, items and stimuli on the same scale continuum.

6. Implications

The major outcome from this empirical study is, perhaps, the potential of incorporating new variables in calibrated structures for the measurement of an underlying attribute of a product. A practical value from the inclusion of further variables is the construction of an item bank. As a consequence, analysts can develop a bespoken structure with additional items without losing the properties of the core of the original, off-the-shelf calibrated scale to make whatever general comparisons they require. We envisage, therefore, that future applications could incorporate computer-aided assessment of the effects of different characteristics of physical components and product features on persons' attitude.

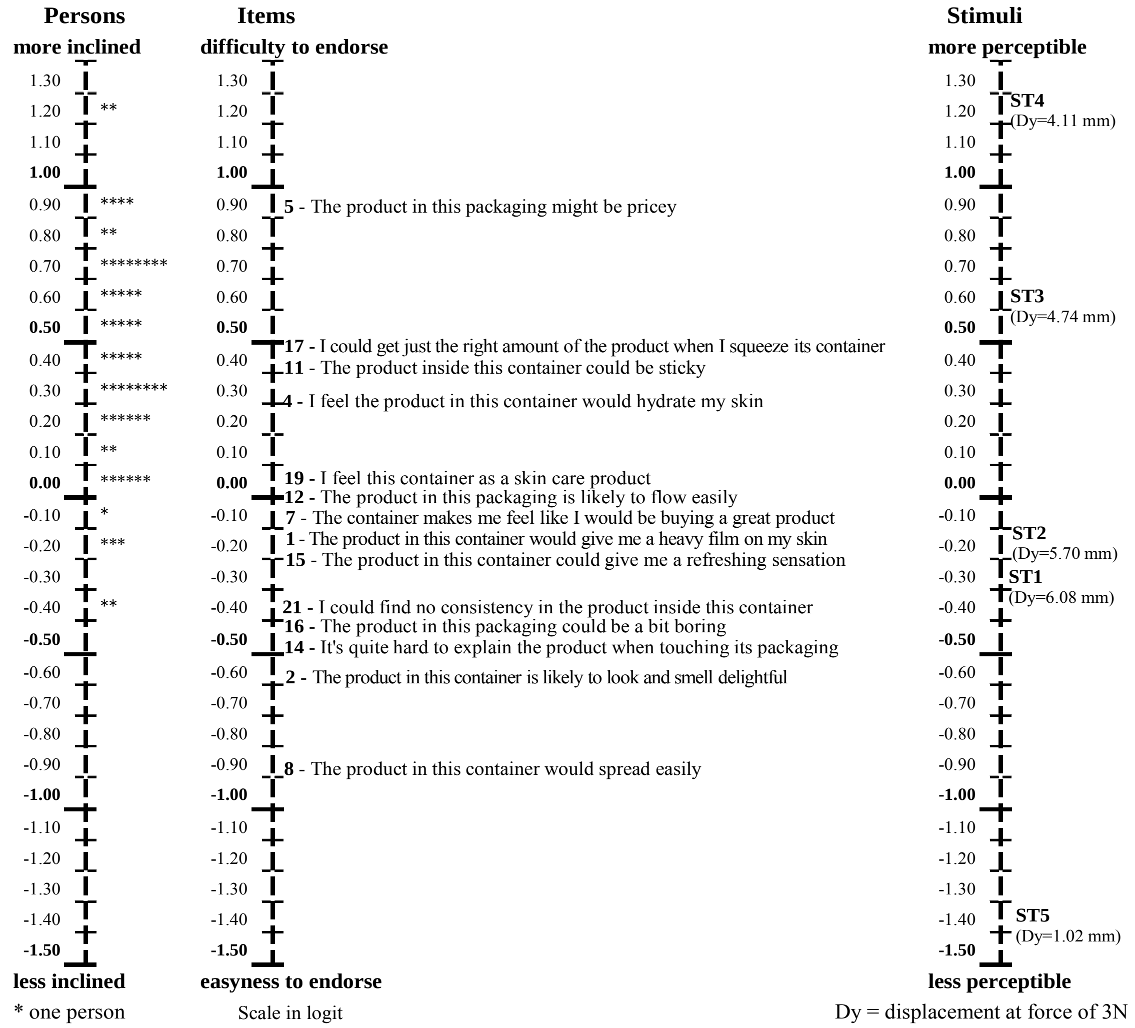

Figure 2. Rasch-calibrated metric including with 14 items for the impression of a moisturizer cream associated with the compliance of a range of product containers.

References

[1] Andrich D 1988 *Rasch models for measurement* No. 68 (London: Sage Publications)
[2] Camargo F R and Henson B 2011 *J. Design. Res.* **9** pp 360 – 75
[3] Rossi G B 2007 *Measurement* **40** pp 545 – 62
[4] Rasch G 1960, 1980 *Probabilistic models for some intelligence and attainment tests* (Chicago: University of Chicago Press)
[5] Andrich D 1978 *Appl. Psych. Meas.* **2** pp 581 – 594
[6] Masters G N 1982 *Psychometrika*, **47** pp 149 – 174
[7] Camargo F R and Henson B 2012. Invariant comparisons in affective design *Advances in Affective and Pleasurable Design* ed Y G Ji (Boca Raton: CRC Press) pp 490 – 499
[8] Linacre J M 1989 *Many-facet Rasch measurement* (Chicago: MESA Press)
[9] Pallant J F and Tennant A 2007. *Brit. J. Clin. Psychol.*, **46** pp 1 – 18
[10] Shao F, Chen X, Barnes C J and Henson B 2010. *J. Eng. Med.*, **224** pp 97 – 105
[11] Andrich D, Sheridan B, Lyne A and Luo G 2012. *RUMM laboratory*. Perth.