Eliciting User Experience through Rasch-Calibrated Metrics for Latent Variables

Fabio R Camargo

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

ABSTRACT

Measurement of affective user experience is not straightforward as it is typically to measure the physical properties of the elements that aggregate a product For this purpose, it is necessary to develop metrics associating observed user experience in the real world with a relevant latent (i.e., unobserved) attribute of the product. However, metrics for latent variables can be undermined by misinterpretation, ambiguity, unfamiliarity, bias, redundancy and multidimensionality. For this reason, anomalies in data ought to be investigated through a robust measurement theory to determine to what extent they corrupt quantitative properties. This paper shows that Rasch measurement theory, which embraces a family of probabilistic models, provides procedures referred to as calibration to test the hypothesis that the metric fulfils measurement principles. As a result, linear scales of affective user experience can be aligned to physical properties of products, allowing generalization of comparisons beyond the particular sample under which a particular product experience was observed.

Keywords: User experience, affective engineering, Rasch model, measurement.

INTRODUCTION

In this paper it is introduced an approach to elicit affective user experience through measurement. Quantifiable properties of scales have been a bottleneck in the emerging area of affective engineering that seeks to integrate established engineering topics with affective user experience. The reason is that affective responses to products are frequently idiosyncratic, culturally located in the consumer's values and dependent on the influence of social groups. As a consequence, measurement of user's experience cannot be considered a trivial matter.

Eliciting users' experience to a product is not straightforward as it is typically to measure the physical properties of the elements that aggregate a product. The measurement of a physical property is associated with a magnitude, i.e., a relevant kind of quantity that can be expressed as a number and a reference or unit (VIM, 2012). Quantitative properties contrast with qualitative characteristics that cannot be represented through different degrees. A brick, for example, is not more brick than any other one. In this paper, affective experience to products is taken as a phenomenon that can reasonably be interpreted not only qualitatively but also quantitatively through users' responses to physical stimuli.

However, affective responses are related to an underlying property of the product, called an attribute, which exists solely as an element of a concept or premise. For example, one can conceptualise that posture while driving is associated with driver's comfort. Although it is possible to measure the driver's posture, a measure of the underlying

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2109-8

attribute referred to as comfort cannot directly be obtained. The attribute can in this case be related to posture because users express some degree of positive or negative response to their experience associated with a particular definition of comfort with certain consistency. Comfort is therefore interpreted as a latent variable; i.e., it is inferred from the manifested responses. The experience, which is expressed in degrees, can be used as measures only if it meets quantitative properties.

One of the benefits from a measurement approach is that its mathematical models reduce the dimensionality of data and consequently the complexity of the hypothetical construct representing the relevant underlying attribute of a product. In this paper some measurement assumptions and the application of the Rasch model (RM) in the domain of affective engineering are discussed. The purpose is to transform statistically discrete observations into a quantitative structure, which is called a metric. The paper shows that the applications of Rasch measurement theory (RMT), which embraces a family of probabilistic models, fulfil measurement principles (Andrich 1988; Bond and Fox 2007). As a result, linear scales of affective user experience can be aligned to physical properties of products (Camargo and Henson, 2012a).

PREPARING THE GROUND FOR MEASUREMENT

The concept of measurement is associated with the properties of metrics established for summarising the users' experience obtained from their affective responses to physical elements of products. The construction of metrics is a technique of relating users' experience in the real world (called manifest or observed) to the attribute of a product (called latent or unobserved).

One approach for establishing metrics of affective responses is to present to a number of respondents adjectives or statements as stimuli related to the underlying characteristic of the product that an analyst wants to know about. Those adjectives or statements will henceforth be called **items**, a term that suits in different domains of application. Additionally, physical stimuli using variations of one or more elements of a product are presented to respondents. After interacting with the product, users give their ratings on a five or seven-option scale embodied in self-report questionnaires. The scale contains the string of items that represent the underlying characteristic and will be replicated for each variation of the relevant physical element. Although the approach for quantitative measurement can be applied to alternative techniques of data collection, such as cognitive performance using a product, interviews and survey with free responses, they will not be discussed in this paper.

To elicit the user's experience it is expected that the string of items will depend to some extent on the degree of a person's inclination to endorse the underlying characteristic of the product. It is also expected that the degree of endorsement varies according to each individual although groups of similar degree of endorsement can arise for a particular physical stimulus. However, there is a degree of uncertainty in the observations. To deal with uncertainty, a response x to a variable X will be observed as a probability. Adapting the example by Andrich (1988), the probability can be represented as a function f of the attitude of a person n, characterised by the variable B_n, and the affective value of item i, characterised by the variable D_i, such that $\Pr\{X=x\}= f(B_n, D_i)$.

Setting Independent Items

Items are in many cases originated from observations of the users' interaction with a product, interviews, search in relevant literature and advice from experts (Nagamachi, 2011, Barnes and Lillford, 2009). However, items can be source of disturbances in data. Disturbances can generate unwanted variance, preventing the data from measurement conditions. Misinterpretation, ambiguity and unfamiliarity can, for example, result in inexpressive responses or simply guess. Another source of concern is the redundancy between items, which can artificially weight particular responses inflating or deflating reliability indices and item discrimination estimates in statistical analysis. To minimise disturbances in a data set Barnes et al. (2008) have suggested removing from a study items that require additional context to be understood by respondents, contain comparative adjectives, consist of non-gradable adjectives and require a prolonged experience. Although those procedures are undoubtedly useful, they are insufficient to ensure a measurement structure in a data set and, therefore, quantitative assumptions ought to be tested.

One of the assumptions when using a probabilistic approach is that responses are statistically independent. Dependence is identified when a person's response to any item in a scale interferes with his or her response to another item within the same scale. Response dependence can, for example, be found in satisfaction questionnaires where a positive rating of a respondent depends on the responses to the preceding items and where that rating will interfere in the way that the responses on the following items are rated (Marais and Andrich, 2008). Dependent items can mislead inferences or decision made on account of means and standard deviations (Smith, 2005).

Establishing a Frame of Reference

One method to elicit users' experience is simply to observe persons' manifestations in natural circumstances. However, this method can be time-consuming and subject to disturbances as a consequence of uncontrolled conditions. On the other hand, controlled conditions can undermine validity by removing the test from a naturally occurring circumstance. To minimise such an effect a metric has to be constructed to ensure consistency and replication. The items in the metric should carefully be designed to provide sufficient information in relation to a particular affective user experience for a given purpose. As a consequence, the items and the users' responses can be characterised as a function of one another.

The concept is similar to a regular thermometer. A measure of temperature can be taken on the basis of the expansion of mercury under some degree of heat. In that case, heat and expansion of the liquid are characterised as a function of one another. Thus, there are two classes of elements involved that define temperature even though many other factors could be included in the system. This simplification exists for practical purposes, establishing a two-way frame of reference, heat and expansion of a body.

In terms of users' experience, one of the entities that form the frame of reference is the sample of persons that interact with the relevant physical elements. Another entity is the string of items used to assess the underlying characteristic of the product (Table 1). In addition, an extension of the two-way frame of reference is necessary for the tests, in which includes the objects used as physical stimuli (Table 2).

Table 1 – Two-way frame of reference

	Items			
Person	Item I_1	Item I_2	…	Item I_i
P_1	x_{111}	x_{121}	…	x_{1i1}
P_2	x_{211}	x_{221}	…	x_{2i1}
⋮	⋮	⋮		⋮
P_n	x_{n11}	x_{n21}	…	$x_{\mathrm{ni}1}$

Table 2 - Frame of reference for eliciting a relevant affective user experience

	Item I_1				***Item I_i***			
Person	Stimulus 1	Stimulus 2	…	Stimulus S	Stimulus 1	Stimulus 2	…	Stimulus S
P_1	x_{111}	x_{112}	…	x_{11S}	x_{1i1}	x_{1i2}	…	$x_{1\mathrm{iS}}$
P_2	x_{211}	x_{212}	…	x_{21S}	x_{2i1}	x_{2i2}	…	$x_{2\mathrm{iS}}$
⋮	⋮	⋮		⋮	⋮	⋮		⋮
P_n	x_{n11}	x_{n12}	…	x_{n1S}	$x_{\mathrm{ni}1}$	$x_{\mathrm{ni}2}$	…	x_{niS}

Functioning of a Metric

To meet quantitative conditions a metric that elicits just one single underlying characteristic of the product is assumed. This is similar to obtain an object's length using a measure tape. If one wants to know about the object's weight, then another measurement instrument would be necessary. However, it is not untypical in the domain the development of scales intended to measure multiple characteristics. The main concept of the approach in this paper is to meet conditions for reliable interpretations on numbers. For instance, a unidimensional scale can allow making comparisons using differences in degree (Andrich, 1988). Therefore, if an analyst wants to know about different underlying characteristics, then a number of metrics, each measuring one single attribute, will be necessary.

The functioning of a metric is subject to the conditions of the administration of the tests as well, which need to be understood and specified. This specification is also necessary in more classical statistical approaches to testing. In that case, persons are referred to in terms of populations. Based on theoretical grounds, classical approaches require that a significant random sample be effectively taken from the population. It is noteworthy that if different groups are considered in a test, for example males and females, and a comparison of their means is expected, then the metric should show that the comparison is invariant with regard to any item. In other words, the persons' interpretation to an item ought to be the same for both groups to yield a solid basis for inference.

Although less familiar than the classical statistical approaches, Camargo and Henson (2011) have demonstrated that the RM can succeed in transforming observed data from affective user experience into consistent metrics that allow algebraic operations and consistent comparisons. Following the Rasch modelling approach, persons' responses are converted into objective measures by calibrating a scale with targeted items established as a yardstick.

A PROBABILISTIC APPROACH: THE RASCH MODEL

The Rasch model, named after the Danish mathematician Georg Rasch, is a probabilistic approach that obtains parameters of persons and items separately. This property allows the design of a range of items in different degrees of difficulty of endorsement on a scale and the distinction between individuals of a sample in different levels of attitude for each item.

The model's procedures test the hypothesis of that the observations meet the necessary assumptions for validating the quantitative structure of the data in hands (Andrich, 1988). Such procedures are denoted calibration, a term coined by Wright and Panchapakesan (1969), referring to measurement scales that are independent of the sample of persons used to estimate parameters of items and independent of the set of items used to obtain scale scores.

Another property of the model is that the raw scores are sufficient statistics to obtain independent measurement parameters. This property distinguishes the model from other probabilistic models and classical statistical approaches. In Rasch modelling, the observed responses are converted into frequencies. These frequencies are then transformed into locations of persons and items on the linear continuum of the metric. Person and item estimates are preliminarily obtained according to a rating scale (Andrich, 1978) or partial credit parameterization (Masters, 1982) and then compared with the observations. The estimates are then revised and new estimates are computed. This process of iteration is carried out until the changes of the estimates are smaller than a stopping rule controlled by a convergence criterion. After the estimates have been made, the data are evaluated to determine the extent to which they fit the model. Most of the estimation procedures are based on the method of maximum likelihood (Fisher, 1922). The estimates obtained from this method point to the values of parameters which maximize the likelihood that the observed data would have generated.

Thus, differently of the classical approach that tries to accommodate a model to the data, the procedures of the Rasch model examine how well the data fit together and cooperate to define the underlying characteristic of the experience being measured. Therefore, the assessment of dimensionality within the context of Rasch measurement theory uses discrepancies between the observed responses and the expected responses by the model. Although a variety of psychological processes are involved when responding to a set of items, it is understood that each item is affected by the same processes and in the same manner. That is, if data fit the model, items are said to be part of a unidimensional structure with quantitative properties (Smith, 2002).

Categories and Thresholds

One of the main factors to achieve a quantitative property in a scale is associated with the order of the response options. It is typical in self-report questionnaires for capturing affective user experience that the direction of responses options is established arbitrarily although an order is implied. For example, an analyst could determine that a mark would be given from 0, which represents the lowest level of endorsement of an item, to 7 the highest level, or from a mark of strongly disagree, representing the lowest level of endorsement of a statement, to strongly agree, representing the highest level. Some analysts also use scales with negative numbers, such as -3, -2, -1, 0, 1, 2, 3. In any case, an order is established to express the users' endorsement level in some way operationally. However, in many studies the scale does not hold the scoring function specifications established by an analyst as a consequence of people's inconsistent use of the response options. The Rasch model procedures examine the consistency of response options using the concept of thresholds.

Thresholds are the transitions between two response options. Conceptually, if the response patterns are consistent, each threshold specifies a point where the probability of a response is equally likely (Andrich, 1978). In a scale with response options from 0 to 3, for example, there are three thresholds. The first threshold is the point between options 0 and 1 where the probability is 50%. The second threshold is the point where the likelihood of response either 1 or 2 is the same. The third threshold is similarly established between options 2 and 3. Each of the thresholds qualifies the average location, or difficulty, of the item on the linear continuum. Analysis of the transitions between categories can be interpreted as though there was an independent response for each of the thresholds. This allows identifying potential problems with the empirical order of categories.

Item Map for Scales with Multiple Response Options

Item map is a representation of persons and items locations on the linear continuum. Once an analysis output has been examined, a general overview of the location values of the items thresholds and persons' inclination of endorsement can be displayed through the item map.

Take Figure 1 as an example. It represents the item map from a study on specialness of wrapped confectionery (Camargo and Henson, 2011). In that study, more than 300 people gave their ratings based on Likert-style, five-point scales to 24 statements contained in a self-report questionnaire. After calibration, twelve items were removed from the preliminary pool as a consequence of anomalies in the data originating from misinterpretation, ambiguity, unfamiliarity or redundancy. The item map in Figure 1 refers to a scale of specialness for one piece of confectionery where the linear continuum has a range from -4 to +4 in logits (unit in Rasch-calibrated scales). The continuum range was produced from estimates calculated through the Rasch-dedicated software package RUMM2030® (Andrich et al., 2012). At the left side of the continuum person locations are, in this example, identified through groups of respondents with similar locations within a range in logits. Locations on the top indicate people more inclined to endorse the attribute specialness than locations at the bottom. At the right side of the continuum the threshold for specific items is indicated by the appropriate suffix appended to the item label, such as the indication 17.4 representing Threshold 4 of Item 17. Locations on the top of the continuum represent item thresholds that hold more difficulty for endorsement than item thresholds at the bottom, which are easier to endorse. The term difficulty of an item is associated with the number of people who used a particular threshold. That is, if relatively fewer people used a threshold of an item, then that threshold is said to hold more difficulty for being endorsed.

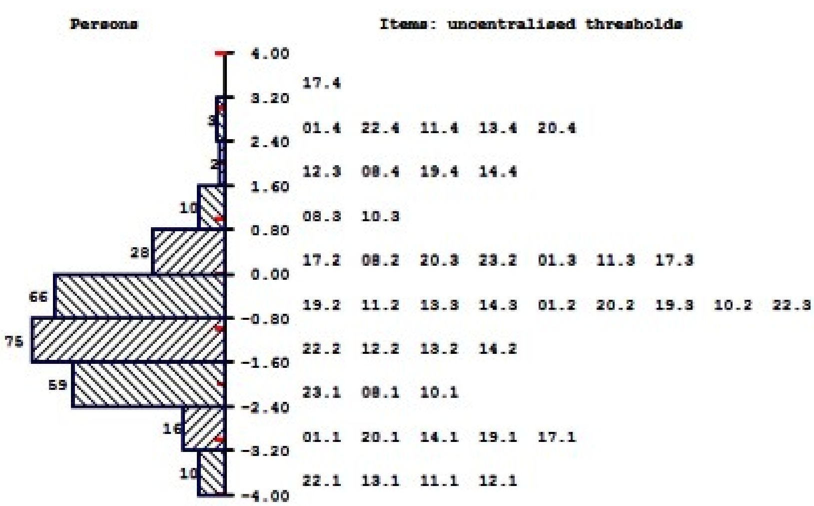

Figure 1 – Item map for eliciting the attribute specialness of one piece of wrapped confectionary (Camargo and Henson, 2011)

Interpretation of a Rasch-calibrated Metric for Affective User Experience

Applications of the RM for constructing metrics of affective user experience require a frame of reference with at least three elements; i.e., persons, items and physical stimuli. In this framework each observation is the result of an interaction among those elements, which are modeled to operate independently. Linacre (1989) conceptualized these interacting elements as facets, developing a derivation of the RM called the many-facet Rasch model (MFRM). The Linacre's model has been adapted by Camargo and Henson (2013) for applications in affective engineering. In the applications, Rasch-calibrated metrics attains interpretation when the difference between persons as well as between items is established by the distance between different locations.

An example of a Rasch-calibrated metric for affective user experience is taken from a study on squeezable packaging of five everyday products (Camargo and Henson, 2012b). The study aimed to establish a range of compliance for the products' containers that could give an impression of a moisturizer cream. In the study, 120 participants touched the containers without seeing the products. They gave their ratings of a preliminary pool of 16 items using Likert-style, five-point scales embodied in computer-based self-report questionnaires. Five items were removed after calibration. Afterwards, three more items were incorporated to metric through a re-calibration using 60 more participants (Camargo and Henson, 2013) (Figure 2).

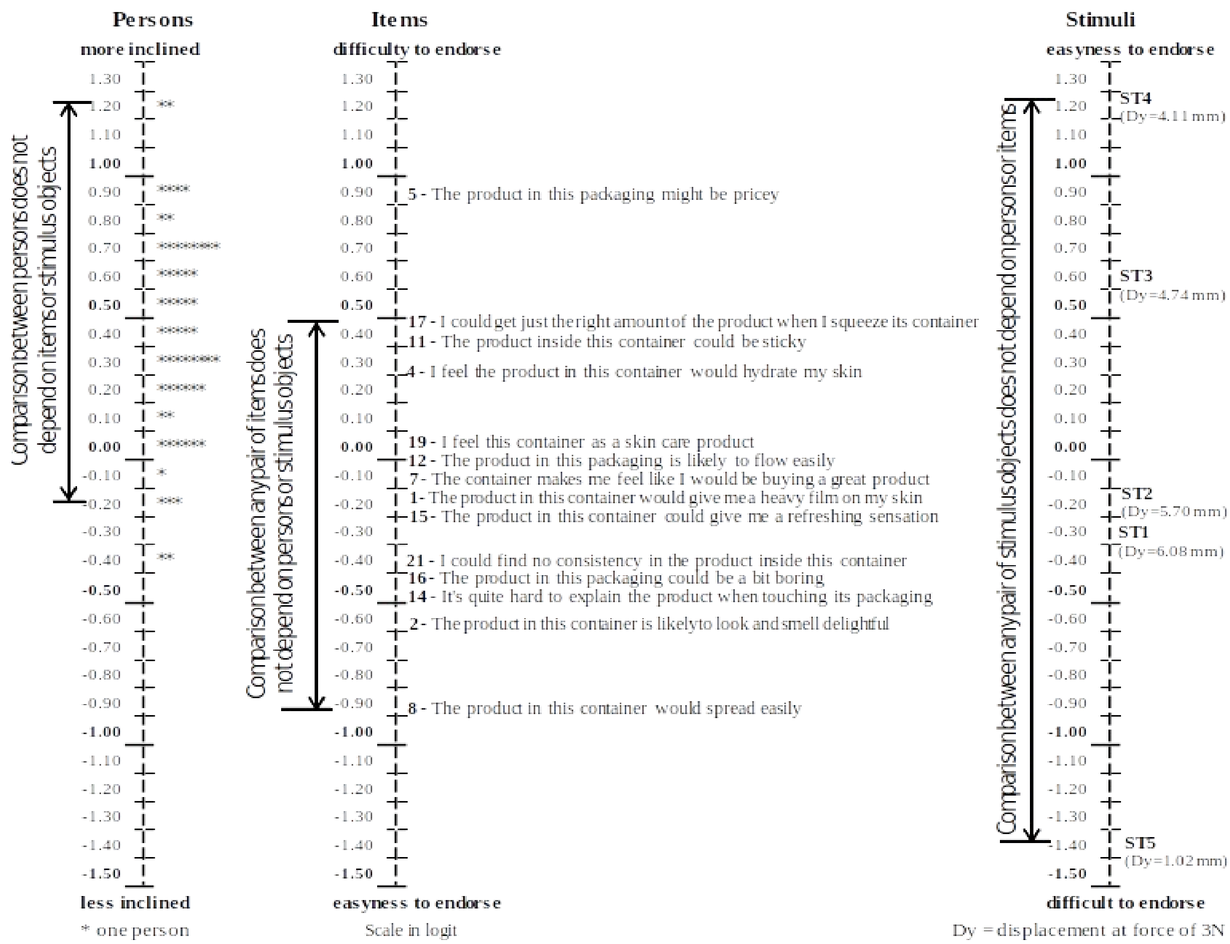

Figure 2 – Example of a metric for affective user experience from a study on packaging

In the metric all facets are on the same linear continuum. The person parameter indicated in Column Persons represents the inclination of endorsement to any item and any stimulus object. That is, the more the readiness, the higher the probability of affirming an item. The item parameter in Column Items indicates the difficulty of endorsement. An easier item is endorsed by relatively more respondents than a more difficult item. That is, the easier the item, the more likely it will be affirmed. The stimulus parameter in Column Stimuli indicates the easiness of endorsement for the relevant attribute of the product. The more the attribute is fulfilled by the stimulus, the more likely it will be endorsed. The interpretation of probability of a response is obtained through a function of the distance amongst a person, an item and a stimulus object. The distance is established by an interval property referred to as log-odds unit or logit. That is, the changes of level in the metric have a constant interval established by a factor of 2.718281…, which is the base of natural logarithm (Linacre and Wright, 1989).

One important characteristic of a Rasch-calibrated metric is the interpretation based on invariant comparisons. Note that the comparison between any two persons does not depend on any item in the metric and any stimulus object either. By a symmetric argument, the comparison of difficulty between any pair of items is independent from any person and any stimulus object. Similarly, the comparison of physical parameters between any pair of stimulus objects depends on neither the persons nor any item in the scale.

The metric allows also the observation of physical parameters associated with affective responses. In the example, the physical parameter compliance, which was obtained from measurement of the displacement of the containers' surfaces at force of 3N, is associated with the location of each stimulus, which was obtained from persons' responses. It is noteworthy that there is a particular range of containers' compliance around Stimulus 4 and Stimulus 3 where the persons' endorsement is higher for a moisturizer cream (Camargo and Henson, 2012b). The metric allow the quantification of the differences between Stimuli 1 and 2, and Stimuli 3 and 4. This kind of information is a valuable resource to establish physical parameters when accounting for affective user experience in product engineering.

IMPLICATIONS

Adding Value to Products

Improvement of existing features has been a common practice to add value to products. Values can be associated with the product experience of users and the interpretation of them. Product experience can fluctuate according to social and cultural oscillations, technological advancements and contextual conditions of each industrial application. Such fluctuations, therefore, produce data from affective responses that could not be statistically stable for time enough and with different groups of persons to be utilisable for generalisation of the outcomes. Nevertheless, some works in affective user experience seem to reside in a false sense of objective interpretation when using statistical reasoning, making hasty generalisations. On the other hand, in many situations a latent regression model based on RMT will make more intelligible the associations between variables (Christensen, 2006) allowing afterwards the application of scores transformed by the RM into a diversity of statistical tools.

The RM adds mechanisms in the process to elicit affective user experience that allow the validation of the structure of data from peoples's responses under certain empirical conditions. Such conditions can be seen through the works of Campbel (1920, 1928), Luce and Tukey (1964) and Krants et al. (1971), associated with the structure of numbers under corresponding algebraic operations. The most important perhaps, the RM can mathematically support the interpretation of the test scores and, as a consequence, to support the implications that the interpretation entails. On the contrary, if just the sum of scores was taken without any validation, the results would contain a high level of imprecision as an effect of systematic errors, and regression coefficients would likely be attenuated. Thus, in affective engineering the RM can find the best number of independent variables in an instrument, maintaining the quality of measurement and avoiding problems with regard to the adverse effects of short instruments and small samples when using classical methods.

Implementation of the RM

A practical value from metrics applied to elicit affective user experience is the potential integration and alignment with different stages of systematic or quasi-systematic product development processes. In most of the cases measurement of affective user experience has basically been a data-driven process based on a very general level under the assumption of a quantitative attribute, undergoing the influence of sampling and systematic biases. As a consequence, the measurement outcomes are difficult to validate empirically and exploit to the different stages of the product development process. However, if evidence of the existence of the affective attribute as a quantitative latent variable is obtained, then it is possible to implement metrics that can contribute to integrate different stakeholders' expectations, resources and decisions associated with consumers' responses to products throughout the process. Although Table 3 is not an exhaustive list of opportunities, it is possible to envisage many of the potential implementations of the RM in a product development process.

Table 3 - Potential opportunities of implementation of the Rasch model in product development processes

Opportunity during a product development process	**Implementation options**
Project scope	Definition of a theory-driven coverage assessment for human-centred design.
Physical requirements	Measurement of latent responses to physical characteristics.
Performance requirements	Measurement and adjustment of the relationship between affective responses and performance.
Usability requirements	Measurement of latent variables in human-factors design.
Interface definition	Measurement and improvement of different interfaces and derivatives with regard to affective responses.
Labelling	Measurement of perception of meaning and comprehension of information.
Inter-changeability	Measurement of attitude toward the integration with other products or accessories of the same family.
Packaging	Measurement of attractiveness and relationship with the product.
Quality management	Measurement of perceived quality.
Reuse and refurbishment	Measurement of attitude toward reusable products.
Affective attenuation	Measurement of the level of affective attachment to a product or components of the product throughout its life-cycle.
Reference documentation	Traceability of calibration of scales for affective responses and affective requirements based on objective measures.
Decommissioning	Measurement of affective responses to premature decommissioning or disposal.
Installation and set-up	Measurement of the level of users' expectation with regard to temporary interruption of services, downtime or difficulty of set-up.

CONCLUSIONS

In this paper a novel approach to elicit affective user experience based on Rasch measurement theory has been presented. It is assumed that in the process to interpret affective user experience, data ought to be evaluated not only by qualitative statistical approaches but also through measurement criteria. The approach's hypothesis is that people's affective responses are related to an underlying property of the product. If evidence of the existence of the affective user experience as a quantitative variable is obtained, then it is possible to construct metrics to associate manifested users' experience in the real world with a latent property of a product.

A metric of affective user experience, therefore, imposes particular requirements on data quality to achieve a constant unit of measurement and invariance across comparisons between the different elements that consolidate the scale. However, anomalies in data can prevent themselves from measurement conditions. The Rasch model provides procedures referred to as calibration to validate the quantitative structure of the data within a determined frame of reference. In the particular case of affective user experience the frame of reference embodies three elements; i.e., persons, independent variables referred to as items and stimulus objects. Within a relevant frame of reference, the Rasch model examines how well the data fit together and cooperate to define the underlying characteristic of the experience being measured, contrasting with typical statistical approaches that try to accommodate a model to the data. The Rasch model requires meeting the key assumptions of independence of the variables established as a yardstick for measurement and unidimensionality of metric with regard to the underlying attribute of interest. Therefore, if data fit the model, the measures will attain interpretation through the comparison between persons as well as between items established by the distance between different locations on a linear, interval continuum.

Based on previous research it is possible to envisage that in many situations a model based on Rasch measurement theory will make more intelligible the associations between variables, allowing afterwards the application of scores transformed by the Rasch model into a diversity of statistical tools. Furthermore, the approach introduced in this paper can overcome some difficulties from eliciting affective user experience through a data-driven process that undergoes the influence of sampling and systematic biases. The theory-based approach using the Rasch model can in many cases validate the data empirically and exploit the results to the different stages of the product development process.

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. (1988). *Rasch models for measurement. Sage university papers series on quantitative applications in the social sciences.* No. 68. London: Sage.

Andrich, D., Sheridan, B. E., & Luo, G. (2012). *RUMM2030: Rasch unidimensional models for measurement*. Perth, Australia: RUMM Laboratory.

Barnes, C. & Lillford, S. (2009). Decision support for the design of affective products. *Journal of Engineering Design*, *20*(5), 477 — 492.

Barnes, C., Childs, T., Henson, B. & Lillford, S. (2008). Kansei engineering toolkit for the packaging industry. *The TQM Journal*, *20*(4), 372 – 388.

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch model: fundamental Measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Camargo, F.R. & Henson, B. (2013). Aligning physical elements with persons' attitude: an approach using Rasch measurement theory. *Journal of Physics: Conference Series*, *459*(012009), 1 – 6.

Camargo, F.R. & Henson, B. (2011). Measuring affective responses for human-oriented product design using the Rasch model. *Journal of Design Research*, *9*(4), 360 – 375.

Camargo, F.R. & Henson, B. (2012a). 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. & Henson, B. (2012b). Invariant comparisons in affective design. In: Y.G. Ji (Ed.). *Advances in affective and pleasurable design* (pp.490 – 499). Boca Raton: CRC Press.

Campbell, N. R. (1920). *Physics. the elements*. Cambridge: The University Press

Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. London: Longmans, Green.

Christensen, K.B. (2006). From Rasch Scores to Regression. *Journal of Applied Measurement*, *7*(2), 184 – 191.

Fisher, R. A. (1922). On mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, (A)* 222, 309 – 368.

Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement*. Vol. 1. New York: Academic Press.

Linacre, J.M. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.

Luce, R.D. & Tukey, J.W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1 – 27.

Marais, I. & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200 – 215.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149 – 174.

Nagamachi, M. (Ed). (2011). *Kansei/affective engineering*. Boca Raton, FL: CRC Press.

Smith, E.V. (2002). Understanding the Rasch model: 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.

Smith, E.V. (2005). Effect of item redundancy on Rasch item and person estimates. *Journal of Applied Measurement*, *6*(2), 147 – 163.

Linacre J.M. & Wright B.D. (1989). The "Length" of a Logit. *Rasch Measurement Transactions*, *3*(2), 54-55.

VIM - International vocabulary of metrology (2012). *JCGM 200:2012 - Basic and general concepts and associated terms* (3rd ed.) Retrieved from www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

Wright, B. & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, *29*(1), 23 – 48.