Int. J. Product Development, Vol. X, No. Y, XXXX

Resonance testing: an industry approach for experiential concept evaluation

Lassi A. Liikkanen*

Helsinki Institute for Information Technology HIIT,
Aalto University,
PO Box 15600, FI-00076 AALTO, Finland
Email: Lassi.Liikkanen@HIIT.fi
*Corresponding author

Heather Reavey

Continuum,
1220 Washington Street, West Newton, MA 02465, USA
Email: HReavey@continuuminnovation.com

Abstract: This paper describes a professional practice in user-centred product concept design that is embedded in a method called resonance testing. It is a nimble method used and developed in industry to ensure that desired properties of design are communicated to the user through the design. It tests product concepts for emotional and functional design attributes such as personal needs, believability, and differentiation. In resonance testing, the users of a specified segment experience design artefacts of variable abstraction levels to see how they perceive the qualities of a concept and how it matches their preferences and expectations. We find that literature lacks both effective user feedback solutions for early product decisions and discussion of the known challenges for doing that. The paper describes how resonance testing generates qualitative insights, builds confidence in new concepts and helps designers to develop the right concept for further development. We present two cases studies of utilising this method.

Keywords: concept design; evaluation methods; design practice; decision making; user experience; experience design; concept evaluation; resonance scales; product development; qualitative research.

Reference to this paper should be made as follows: Liikkanen, L.A. and Reavey, H. (XXXX) 'Resonance testing: an industry approach for experiential concept evaluation', *Int. J. Product Development*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Lassi A. Liikkanen is an Interdisciplinary Researcher and Design Practitioner. His PhD in Product Development was awarded by the Aalto University, Finland. His post-doctorate period included an extended research visit to Stanford University and short visits to China and Japan. His main interest lies in design cognition and creative design practices. His authored works appear on over ten international peer-reviewed journals and in numerous conference proceedings, including reviewers' favourite paper from ICED 2013.

Heather Reavey leads Practice Innovation at Continuum, designing the strategic trajectory of all practitioners, disciplines and methodologies to ensure that Continuum delivers exceptional results to client partners and maintains its

leadership position in the business of innovation. She has contributed to many of Continuum's most acclaimed projects over the past 15 years, helping to launch new strategies and offerings for Procter & Gamble, Insulet, American Express, United Healthcare, Tetra Pak, Kraft and BBVA. The experience she is gained has equipped her to anticipate how changes in people and cultures will impact business, and to develop breakthrough ideas that will win in the changing marketplace. She is a graduate of Syracuse University's Industrial and Interaction Design program, with a focus on visual storytelling. She is also a Professor of Strategy and Innovation at Massachusetts College of Art and Design.

1 Introduction

The scope of design has expanded over recent decades. Designers are required to design not only products and services, but also focus on experiences, values and meanings (Krippendorff, 2006; Redström, 2006). The importance of the users' perspective in design has also steadily increased. For instance, design thinking has been advocated as a new paradigm for developing organisational capacity for innovation (Brown, 2009). This movement has audibly promoted the methods and ideals of human-centred design for improving business. At the same time, some have refuted the strong emphasis of users' opinions in creating radically new innovations because regular people cannot be expected to show much foresight to their future behaviour (Norman, 2010). The question is how can design organisations harvest the benefits of user-centred design while balancing users' and designers' perspectives? The challenge we address specifically is how to elicit user reactions when making concept decisions.

We know that companies utilise design and market research methods to improve the success of their new products. Unfortunately, the details of their practice are often considered confidential and remain unpublished, whereas design and evaluation methods from the academia can be unsuited for the abstraction level of the design or otherwise impractical (see e.g. Vermeeren et al., 2010). This paper addresses this gap by introducing a light weight solution for user involvement in concept testing, called *resonance testing*. This procedure has been developed at a global design company called Continuum and has been used in various commercial product development projects.

In brief, the idea of resonance testing (RT) is to simulate by presenting new concepts on different abstraction levels to users to evoke user experiences and evaluate users' perception. It allows the designers to see if concepts can convey the intended meanings and user experience (UX) targets in the first exposure and use episodes. RT provides qualitative feedback for the designers to see if the values they intended to embed in the design are perceived by the user as intended, and whether the new experience that they have designed is better than existing experiences that the user may have as reference. Understanding how people read, understand, and evaluate new ideas helps designers and innovation teams to iterate and improve designs, and make more informed concept decisions. Ultimately, it gives teams confidence to continue to invest in early ideas.

This evaluation method builds upon the elements of card sorting and laddering interviews. It reveals whether the design evokes the visceral responses and perceptions of functionality envisioned by designers; does the product resonate on functional and affective levels; does it generally work and does it work for me. This method acknowledges the requirement of aligning design with other company activities, foremost marketing and

business management. The method fills some gaps identified in the UX evaluation literature (Vermeeren et al., 2010) with its early design applicability, testing of multiple people together and practicable nature.

In this paper, we aim at communities interested in practices of designing for and measuring product experiences. With this document of an industry practice, we wish to allow scholars and product designers to get inspired by the elements of the method (see Woolrych et al., 2011), and get an insight onto how practices (Liikkanen et al., 2011) instead of strict methods (see Laakso and Liikkanen, 2012) rule in the professional work. For motivation, we review the existing literature on the methods and goals of concept testing by demonstrating the lack of known, comparable approaches. We introduce many ideas from the literature on interactive systems design, in which the recent conceptual and methodological advances should inform experiential product development as well.

The name resonance testing requires some clarification. In our vernacular, RT, or resonance, is not a concept like UX, but a name for a method. Resonance testing can involve UX targets, thus it can be also considered a UX evaluation method. Because we do not have a fixed definition of resonance, we can select the experiential targets as necessary. The term resonance has been previously used in the design context, but in a slightly different meaning. In interaction design, Hummels (2007) discussed resonant interaction, which she defines as the "*perfect interplay between a person and a product*" (Hummels, 2007, p.21) due to an ideal match between user characteristics or demands, and the properties of the product. This definition likely captures how many designers understand the concept. The notion of resonance in RT is slightly different. We agree that resonance occurs when products are acceptable to personal needs and values, but in RT, we do not expect a perfect match. RT can inform designers the best when it helps test users to verbalise why the concept mismatches their ideals. Finally, the term 'testing' can be misleading. The goal of Resonance Testing is not to crown a 'winning idea.' Instead, ideas are evaluated in relation to each other and competitive products through scales that represent the designer's assertion of unmet needs that will connect with users. The goal is to understand the 'whys' in how users evaluate ideas so the ideas can be tweaked to more effectively resonate with users.

2 Background

The professionally managed design process involves understanding people and their context of use, identifying unmet needs, specifying design goals, and satisfying them through the design of a new experience. This should lead to desirable products that people want to buy and accept into their lives. This section examines the need for setting targets or goals for concept testing. The following sections analyse possible goals for concept testing and introduce existing design evaluation methods.

Design needs meaningful goals and constraints. Recently, emphasis in design rhetoric has changed from technical to human-centred goals (Redström, 2006). Setting goals and determining requirements is challenging. Discussing the problems in software products, Cooper argues that many companies fail at this stage; organisations do not "*describe the reaction we'd like the end-user to have*" (Cooper, 2004, p.42). In other words, the experience the end-user should have with the product is unspecified. Setting goals in terms of *experiences* reflects the recent shift from specifying design goals as instrumental properties to experiential properties of use (Redström, 2006; Hassenzahl et al., 2010;

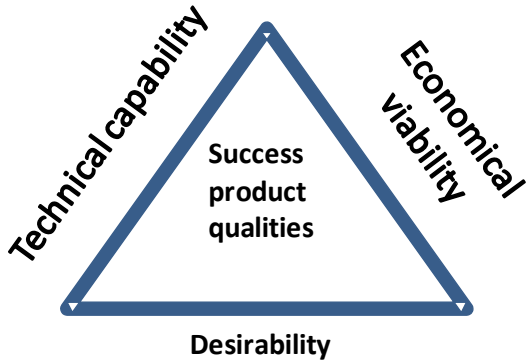
Väänänen-Vainio-Mattila et al., 2008a). The next step after setting goals is to develop the concept demonstrators and evaluate them. This paper will bypass concept concretisation and focus on user evaluation.

The tradition of testing new product ideas with potential users comes from marketing (Wind, 1973; Iuso, 1975), where this activity is known as prelaunch market analysis (Lees and Wright, 2004). According to Moore, "*the purpose of concept testing is to estimate consumers' reactions to a product idea before committing substantial funds to it*" (Moore, 1982, p.279). One of its typical functions is to provide a 'go or no-go' decision for the process (Iuso, 1975). From a design perspective, we must acknowledge that design and evaluation activities are hard to separate (Väänänen-Vainio-Mattila et al., 2008b). Hence, evaluation should serve both an inspirational and research function (Karapanos and Martens, 2008).

3 What to test concepts for

Or in other words, what to design for? Where do design targets come from? Some design and business people share an idea about which factors make a success product. Larry Keeley is cited by Cooper et al. (2013, p.10) as the author of *capability*, *viability* and *desirability* a model for successful products (see Figure 1). Brown (2009) holds a similar view, although he calls capability *feasibility*. Capability refers to the technical feasibility of implementation – what can be reliably delivered. Viability refers to business aspects – whether the idea is sustainable and can make a profit. Desirability reflects the added value, the motivation, for the users to acquire and use the product, explaining why they would want it.

Figure 1 Three factors of a successful product (Brown, 2009; Cooper, 2004)

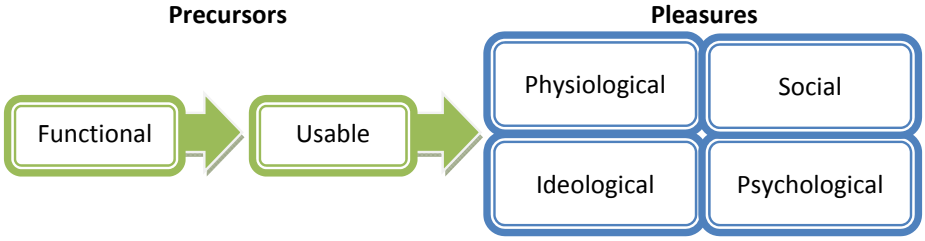


However, desirability, as broadly defined above, does not make a specific enough a goal for concept design or testing. But there is an influential branch of modern design literature that addresses a similar feature, that of *user experience* (Roto et al., 2011). In UX discussions, hedonic qualities of product use (be-goals) are equally important as pragmatic attributes (do-goals) for the design and evaluation of UX (Väänänen-Vainio-Mattila et al., 2008a; Hassenzahl, 2004). Even though designers and scholars still currently mean slightly different things by UX, it seems to be the best candidate for decomposing desirability into design targets.

Resonance testing

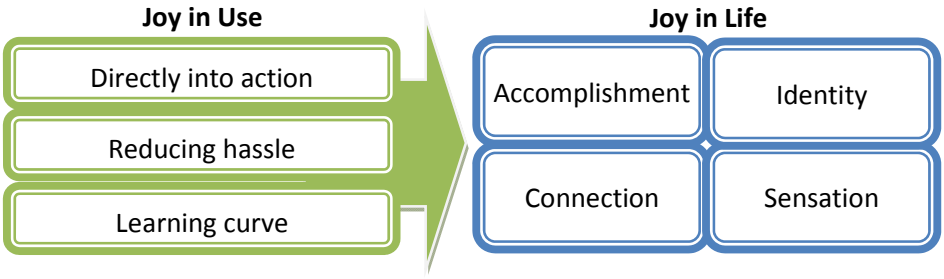
There are slightly different views as to how UX, particularly positive UX, should be divided into more tangible targets. Norman (2004) discussed the emotional side of product experience. His account started from the three levels of human information processing: visceral (immediate reactions, emotion, beauty), behavioural (performance and utility, usability) and reflective (meaning, personal attachment). Another way to look at experiences is to consider distinctively positive experiences. Jordan (2000) identified four types of pleasure derived from products: physio, socio, ideo and psycho pleasure. In this view, pleasure is dependent upon adequate usability, and usability is based on appropriate functionality. According to Jordan, a right mix of pleasures should have preference over usability targets in design.

Figure 2 The four pleasures of Jordan (2000) as prototypes of desirable experiences



A design-inspired model of desirability or 'cool design' was introduced by Holtzblatt (2011). She uses two models of joy to explain how the experience of cool arises from product use, as visualised in Figure 3. *Joy in use* is concerned about instrumental values that are evoked by using the product, which gives rise to *joy in life* that involves four factors which describing how products can make us feel in life.

Figure 3 The emergence of cool according to Holtzblatt (2011). Joy in user mediates the emergence of joy in life



Overall, this section has considered different types of factors, which reflect desirability. Norman, Jordan and Holtzblatt seem to converge that the emergence of desirability is a combination of positive experiences in use, which fulfil different types of user needs – not only instrumental ones. For concept design and testing purposes, it seems necessary that the desirability breakdown must incorporate aesthetic, emotional and individual targets as well instrumental and behavioural ones. Concept design should generate 'experience blueprints' (Brown, 2009) or UX targets (Väätäjä et al., 2012) that state the

desired product's perception in terms of affective and functional qualities evoked. Designers must impute the desired qualities in the concepts, and concept testing should inform how well and why these targets are fulfilled by the designs.

4 How to test concepts

To enable the success of a future product, designers should understand the opinions of respective users. How to achieve this is an old challenge (Dreyfuss, 1955), still without a clear solution. In product development companies, concept testing methods are widely used (Peng and Finn, 2008), even though organisations neither usually disclose nor necessarily know how useful the techniques are. UX evaluation methods have recently emerged in interaction design research. However, scrutinised and public documents describing the test methods used in industry are rare. For instance, the tradition of simulated test marketing, comparable to modern software alpha testing, has been around since 1970s, but has not evoked much academic interest (Wherry, 2006). As a result, commonly described UX assessment methods (Hassenzahl and Tractinsky, 2006; Pifarré and Tomico, 2007; al-Azzawi et al., 2008; Roto et al., 2008; Roto et al., 2009b; Kujala et al., 2011) are predominantly developed and applied in academic studies. An extensive analysis of UX evaluation found that only 14 out of 96 (18%) methods inspected originated from industry (Vermeeren et al., 2010). The same study documented that under one-fourth of the methods were applicable to concept or non-functional prototypes and the most frequent issue with them was applicability.

4.1 Challenges and requirements for testing

In order to usefully test concepts, we must take into account a number of potential confounds and constraints. The former have to do with subjects of the study (humans and their experiences) and the latter with a design process that seeks to involve outsiders. We must acknowledge that UX with a product is not singular but a *temporally continuous relationship*. Recent studies of UX have differentiated the stages of experience (Roto et al., 2011; Kujala et al., 2011) posit that UX begins from users' expectations and accumulates through momentary episodes of interaction into a cumulative UX, following a trend in marketing research (e.g. den Ouden et al., 2006). This means that prior experiences with products and brands create expectations that influence experiences and product appraisals (Raita and Oulasvirta, 2011; Creusen and Schoormans, 2004). Previous research has shown that reading product reviews prior to use can influence user evaluations of their instrumental and non-instrumental qualities (Raita and Oulasvirta, 2011). The lesson is that user opinions are grounded in expectations but must be evoked by experience.

This nature of product UX makes expectation management essential for concept testing. People habituate to products' performance and learn to expect certain value from everyday products (Jordan, 2000). This can build up unrealistic expectations for new products. A study by den Ouden and associates (2006) discovered that users complained about features that they expected to find but that were never included in the product specifications! Obviously their needs were never targeted by the product designers.

Testing must consider user's expectations and product perceptions related to previous experiences. This means involving the intended users of the product, because the

designer and the users hardly ever perceive the design in the same way, no more than users with different expertise levels do (den Ouden et al., 2006). Norman (1988) articulated this in his proposal of user and designer mental models, and Karapanos and Martens (2007) have empirically shown how designers and users have different perceptions of products. Segmentation (Smith, 1956) is thus important for testing, as it means focusing on clusters of people, who are assumed to have similar product's perceptions.

Experience with products exceeds individual concepts. People can also read visual design language. MacDonald (2001) calls this aesthetic intelligence that enables people to perceive a wide range of qualities in products, sometimes subconsciously. The important consequence for products is that there are stable relationships between product appearance and user's perception of instrumental and non-instrumental qualities (MacDonald, 2001). As an example of such, Creusen and Schoormans (2004) present six values people commonly impute from product appearance. These roles are aesthetic, symbolic (personal), ergonomic, functional, categorisation and attention drawing. In their study, aesthetic, symbolic and ergonomic values were most commonly stated as reasons for a product choice.

How do previous experiences influence the testing of novel concepts which users cannot be familiar with? Here, an issue is that there are hardly any feasible methods for testing novel concepts (Jordan, 2000; Vermeeren et al., 2010). A study of product management practices in quickly technology changing markets (Mullins and Sutherland, 1998) discusses the inadequacy of traditional testing and the need to use prototypes during concept design. The difficulties faced by traditional approaches can be explained by the recent psychological findings that people generally dislike new ideas (Mueller et al., 2012). Consequentially, a survey of concept testing methods found that marketing research methods were predominantly used with incrementally new products (Peng and Finn, 2008), not so much with radically new concepts.

In addition to these fundamental questions about user testing, user involvement in industrial design process has some constraints of its own. Contemporary design favours a short iterative design process. This generates several *requirements for feasible concept evaluation*. For evaluation methods to be useful in the professional use, they have to be lightweight, credible, readily deployable, widely applicable to products and prototypes, accessible for different types of users, satisfy multiple stakeholders, produce systematic output (Väänänen-Vainio-Mattila et al., 2008b), fit different design stages, and capture users' diversity (Karapanos and Martens, 2008). The aforementioned review of UX evaluation methods (Vermeeren et al., 2010) pointed out a requirement of practicability and stated that the most of weakness in the reviewed methods were related to practicability. Finally, organisational resources are always scarce and testing that requires substantial investment will be difficult to adopt.

5 Evaluation methods

There are a number of methods available for evaluating concepts. We cannot provide a detailed review due to space constraints but we will point out some important ones from market research and UX evaluation traditions. In market research, concept testing has been usually based on concept statements, which are typically delivered through telephone surveys, or nowadays online panels. These surveys present questions about

purchase intention, believability, uniqueness and problem-solving to be answered on 5- to 7-point scales (see Duke, 1994). From 100 to 200 responses are usually sampled for a quantitative data set (Iuso, 1975; Moore, 1982). Gathering of more qualitative data is uncommon. The survey results from Peng and Finn (2008) showed that only about one-third of the 50 responding organisations had users who meet with the R&D people.

In the wake of the recent interest in UX, a number of new methods for UX evaluation have been proposed (Väänänen-Vainio-Mattila et al., 2008b; Roto et al., 2009a) and a dedicated website enlists 84 such methods.[1] Among the methods that have been used, card sorting is a simple method for understanding categorical user's perception. It involves classification of objects into participant-defined categories based on sorting criteria given by the designer (al-Azzawi et al., 2008) or freely selected as attributes to describe a product (Benedek and Miner, 2002). This approach has been used for UX evaluation (Williams et al., 2004). Cards depicting emotions associated with products were used by Hummels (2007). There are similar methods drawn from psychology, such as semantic differentials (Chuang et al., 2001; Hsu et al., 2000) and the repertory grid technique (Tan and Hunter, 2002; Karapanos and Martens, 2008). Both rely on rating adjectives, such as scoring a flashlight on a 1- to 7-scale between 'bright–dim', but analyse data differently. There are some variants, such as the anticipated experience evaluation, or AXE, method that uses paired images instead of adjectives for evaluating experiences and stimulating discussion about the experience (Gegner and Runonen, 2012). It also involves an elaborate framework for the analysis of the session transcript, which should produce feedback for design.

Interviews and questionnaires can also be used. Laddering interviews have been used in marketing to discover people's values and their impression on the product's meaning (Veludo-de-Oliveira et al., 2006). In laddering interviews, respondents are asked consecutive elaborating questions, typified by the '*Why is that important to you*' question (Reynolds and Gutman, 1988). Variations of laddering, for instance bipolar laddering (Pifarré et al., 2009), have been introduced for UX evaluation. Jordan (2000, pp.130–133) presented a pleasurability questionnaire developed at Philips as a way to assess UX. This test presents experience evaluation goals in a quantitative format, several similar instruments for measuring emotion are listed on the AllAboutUX website.

Finally, some methods assess the association of visual features to product meaning and design targets (Schoormans et al., 2010; Artacho et al., 2009). They involve extensive collection and analysis of user data, and mapping this to experiential features to design. This is similar to the design relational chart method from Kansei engineering, which has been successfully used by Japanese corporations to design consumer goods (Nagamachi, 2002). However, in general, these methods are often focused on design (mapping features to targets) rather than evaluation and are too heavyweight for professional practice by the previous criteria (Väänänen-Vainio-Mattila et al., 2008b; Karapanos and Martens, 2008).

6 Resonance testing

Resonance testing was developed at Continuum LLC (previously Design Continuum, Boston, MA) in the late 1990s. The company found it difficult to test both new-to-the-world and incrementally new concepts. The problem with the latter was that people were too accustomed to existing everyday designs (here, padlocks and window frames) to

articulate their needs and appreciate new, alternative designs. The assessment of radically new products had a similar problem (see Trott, 2001). Concept tests implied that people did not like new ideas, leading to misinformed product decisions, when in fact similar designs from competitors soon fared well in the market. These issues sparked the development of resonance testing, a new method to elicit actionable feedback, understand *user's* perception, and help to refine and select the ideas. Since then, the company has seen considerable gains from using the method and recommends RT for most design projects offered to clients.

RT is a market-oriented but human-driven formative concept evaluation method, suited for rapid, iterative concept development and supporting comparative evaluation. It aims to improve and inform concept decisions in the early development stages. It embeds theoretically grounded ideas in a format that meets the feasibility requirements of resource constrained industry context. It assumes a design process aimed at a defined domain, customer segment and design targets. RT is part of a 'bottom-up' concept design approach, promoting exploration through building, testing and refining. The role of users can be illustrated with a trial metaphor: designers carefully frame and present the evidence, user testifies, but designers make the verdict based on the testimony. RT is organised to help the testimony by providing a setting, which stimulates discussion and helps users to articulate their experiences.

6.1 What to test for in resonance testing?

RT is conducted to ensure that the future product can be competitive in the market and deliver instrumental (usability and functionality) and experiential values. RT holds the design assumption that designers have done research for design and created hypotheses about *who* the users are, which values they hold and *what* they desire. The desired UX targets are shaped into design artefacts, which should already communicate high standards of utility and usability (although these are not UX targets themselves). RT is then used to test whether users' perceptions match the set of non-instrumental and instrumental (problem-solving capacity) design targets. The typical design targets are about UX; non-instrumental and affective attributes, for instance: believability, problem-solving and personal relevance.

6.2 Why are concepts tested with resonance testing?

RT happens during the concept design, so it provides designers feedback to revise the concept. Compared to typical concept test methods in market analysis or usability testing, RT is primarily a qualitative, formative evaluation method that tells designers whether the design is evoking the desired percept. RT focuses on the initial experiences with and expectations about the product, and is appropriate for cross-sectional user studies. It helps designers to make decisions about which alternative, if any, is the most promising for further development. RT provides a proof-of-concept by demonstrating a favourable concept's perception in the target segment. RT can be a milestone (a go or no-go step) in the design process and helps screen ideas and assure the client that the design can deliver the desired qualities better than the competition.

6.3 Applicability

RT has been developed for the needs of a large design consultancy with diverse client challenges. Their offerings that range from opportunity identification to product and service experience to the engineering and deployment of these experiences within a client organisation. RT is used in many types of engagements after designers have envisioned ideas or in early design phases as design embodiments are being explored. RT requires a small team of designers to envision the testing protocol and embody the ideas and competitive products so that users can experience the full set of products and evaluate them by referencing the experience. Resonance testing requires a test space, recruiting efforts to access users within the same customer segment, and time and resources for beta tests to refine the protocol and artefacts for testing.

7 Guidelines for resonance testing

The RT procedure involves a user test set-up similar to usability studies with two specific features: laddering levels of presentation and resonance scales. Laddering levels of presentation refers to how concepts are introduced to participants, laddering from abstract descriptions of ideas, in words, to detailed physical embodiments. Resonance scales are an exercise consisting of a set scales that are anchored by design goals – typically a positive and negative description of an attribute. The user would be asked to place representations of the various ideas in space in relation to the anchors and the other ideas, and describe, in detail, why they placed them where they did. Scale exercises include competitive products to ensure that new experiences are perceived as better than existing experiences. A typical test situation is shown in Figure 4 and described in detail in the following subsections.

Figure 4 Photo from a resonance test session. Designers on the left, the participant explaining her choice on the right. Resonance scale is on the table and cards between represent the concepts assessed by the participant

7.1 Artefacts

The stimuli of the resonance testing are the demonstrators of the concept, called design artefacts. One test can involve several (competing) new ideas, along with competing products. Competing designs are modified to match the fidelity level of the new design and selectively include brand information. This allows benchmarking new concepts among themselves and against competition. The fidelity of the artefacts can be variable across test sessions, but should be fixed within a test (cf. Roto et al., 2009a) so the ideas remain comparable.

The spectrum of artefacts ranges from traditional concept statements and illustrated ideas on paper to illustrated use scenarios, video scenarios, paper prototypes, non-functional prototypes, Wizard-of-Oz prototypes (Klemmer et al., 2000), semi-functional prototypes and finally functional prototypes. The resolution of artefacts is important to ensure quality of feedback – if models are too resolved, there is a risk that users will overanalyse the design of the idea and not the idea itself. For example, they might react to a colour or font decision versus the concept at large. Fully functional prototypes are often unnecessary investment; faking the functionality will suffice to evoke more elaborate users' responses than a written concept statement does. The decision of which artefacts to use is made by the design team. Artefacts are subject to change between RT sessions as unforeseen functional or usability flaws discovered by the users may require rapid revisions to the concepts or their demonstrators between the test sessions.

7.2 Timing

RT can be performed as soon as the concept idea(s) have been developed into artefacts. In projects carried out by continuum, RT has taken place within the first six months of the project. The rule is, the earlier, the better, so commitment undesired designs can be avoided. Continuum's concept design process goes through the steps of alignment, research, analysis and synthesis, envisioning-embodiment (construction), resonance testing and delivery of concept statement and artefacts. This process is not necessarily linear. These steps indicate that RT can conclude the concept development process or provide a reason to step back to envisioning or even research if that is suggested by the results. Completing tests can take from two to three weeks, depending on the number of customer segments and geographical sites involved. Using multiple sites requires more planning and tight scheduling, and some time must be reserved for revising the protocol and iterating the prototypes.

7.3 Participants

Early in the design process, designers define one or more 'attitudinal segments' for which concepts are developed. A segment can refer to both cultural and geographical boundaries (country, age group, gender, class, etc.). The test users are recruited to match the defined segment(s) and expected to respond to the concepts in a similar way. If there is any reason to believe that there are important differences across the market areas, testing should be carried out at different locations (e.g. the east and west coast of the USA) or the world (Europe and Asia). It is necessary to screen participants for segment fit. In practice, this means developing checklists for use of similar products, lifestyle and so forth. Experience with related products is particularly important consideration as it

will greatly influence the test results and also sets requirements for the design of the test (e.g. benchmark products should be familiar for the participants). Screening out subjects who cannot be trusted with business-sensitive information is also a standard procedure.

RT accommodates more than one participant for a test session. It allows pairs of users or small families to participate if the product use or acquisition presumable involves more than one person (e.g. a couple buying a boat). Sample size is set according to the goals of the test. The aim is to prove wrong the hypotheses about the concept (what we believe the artefacts should communicate). A sample of six to seven representative users is considered adequate to require adequate feedback. Although crucial for cost involved, the number of participants is not fixed and to best of our knowledge, there is no scientific basis to lean on either. If there are multiple user segments, the number of participants grows accordingly. Informants can be compensated for their participation, but this should be minimal to avoid a kind of reciprocity that might bias informants' opinions.

7.4 Researchers and setting

Evaluation is carried out by designers trained through apprenticeship. It is recommended that designers who interact with participants are separate from the team that has created the concept to remain as neutral as possible. Researchers must not give any unintended details about the product, justify the design rationale, or otherwise influence user's expectations or perceptions. The original team can do preparations and observe the testing inconspicuously, but should avoid contact with the informants. The same goes for management or client representatives, who can gain insights into and confidence for the concept by overseeing the test, but should never interact with the informants. The RT sessions must be excessively documented. Every action and comment by a user should be traceable via an audio-video recording. Multi camera set-ups are used if possible. In practice, the notes made during the session are the primary data and the video is used for checking and extracting key insights for client presentations. Transcribing video and analysing the recordings has been considered to provide too little return for the resources needed.

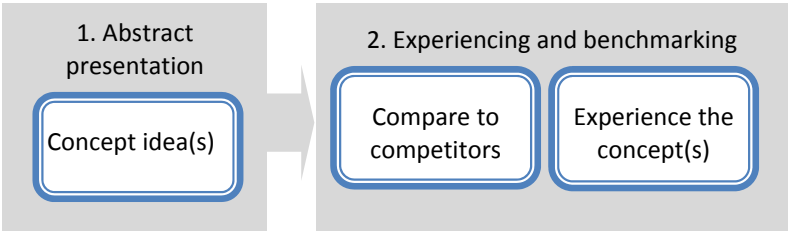
7.5 Procedure

After screening, the participants are invited to a resonance test at the testing site. At occasions, the subject may receive homework related to the product domain, for instance, brand and product awareness questionnaires, to be completed beforehand. The participants are informed about consent, compensation and confidentiality requirements before they arrive. At the testing site, they receive necessary background information and are introduced to the product category and intended use context. The sessions start by an introduction of the participant to the design researchers, one or two of them. This is followed by an open-ended interview that aims to provide lively background information about the participant(s) and their current relationship with the product domain – all relevant context data for the following RT.

The RT procedure consists of introducing the concept step-by-step, from abstract to more concrete artefacts, and eventually letting the user experience the concept through the artefact (see Figure 5). This approach is called *levels of presentation*. The rationale is to let the users experience the artefact on multiple levels of presentation and then probe their perceptions with *resonance scales* (see below) and interview questions. Initially,

users will be exposed only to the plain conceptual idea and later they experience the prototype as well. This makes it possible to discover opportunities for improvement. For example, if the abstract idea is well received but the prototype is rejected, designers must consider whether the prototype can be enhanced or is the idea just unfeasible. Procedure also involves benchmarking the new artefacts against existing market offerings and familiar brands. This is necessary in order to assess the market survival potential of the concept.

Figure 5 The flow of RT procedure, from abstract presentation to concrete experiences with the artefacts



The detailed protocol of the test depends on the fidelity of the artefact. At minimum, the concept should be seen and at best experienced in a realistic context. A written protocol including the resonance scales and interview questions are made to ensure consistency. This is particularly helpful if several researchers are running the study across different sites. RT protocol is also subject to revision between test rounds. RT is sometimes embedded in an extensive user test which includes usability and market research sections.

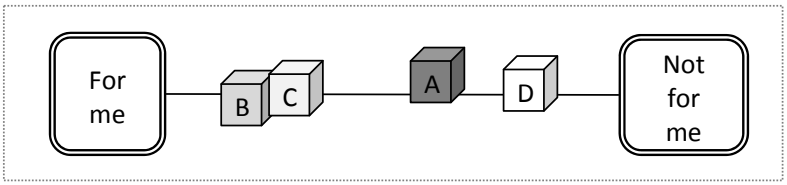
7.6 Resonance scales

After the concept has been introduced, RT examines the concept's perception around the selected design targets, such as believability, impressiveness, sustainability, problem-solving and personal relevance. These are resolved using interviews and interviews stimulated with *resonance scales.* Resonance scales resemble semantic differentials, paired comparison and card sorting. They involve simultaneous comparison of multiple new concepts, and possibly benchmarked products, on one dimension between two scale extremes. These scales include descriptions such as '*intriguing–not intriguing*' (impressiveness scale) '*durable–not*' (sustainability scale) and '*for me–not for me*' (the personal relevance scale). Designers select and create scales based on UX targets and concept-specific needs. Some scales are used regularly, especially the personal relevance scale, as illustrated in the next section.

The participants respond to scales by placing the artefacts, or cards representing them, between the extremes, with physical proximity indicating the standing of each concept (see Figures 4 and 6). As a variation, simplified resonance scales including only one target (e.g. 'best for home use') may be shown and participant picks the best matching example. The results are recorded with a photo. Evaluating multiple concepts simultaneously brings out semantic clustering of the perceived qualities and is a tangible activity, akin to card sorting. Although scales can produce quantitative results through rank ordering, they are primarily used to help to stimulate further interview and find the causes for differences in users' perceptions. After participants make assessments,

designers follow up with interview questions. For instance, after a participant indicates that A seems more durable B, researcher could ask why – similar to a laddering protocol. Other examples of questions are '*Who would use this*', '*How does this perform for X*', '*Why do you believe this product can deliver that?*' and '*Why is this relevant to you?*' Even though the administration of the scales can be quick, usually no more than four scales are used, because they tend to elicit lengthy discussions.

Figure 6 Example of a resonance scale use for comparing four different ideas (boxes A–D) for personal identification



7.7 Test outputs and design inputs

The RT sessions produce several types of data, prominently qualitative resonance scale rankings (photos) and associated interviews explaining users' perception. Qualitative data are used for hypothetic-deductive inference, to either support or falsify designers' expectations about the users' perception of the concept. The analysis is pragmatic and organised according to main variables of the study, prototypes, features, benchmarking or user segments. Qualitative data are used in two ways: (a) summarising findings for each user segment for design input and (b) transcribed excerpts from interview protocols are used to illustrate the concept reception for project stakeholders. In some cases, video clips are used to complement the excerpts. Results can be summarised concept-by-concept, or by going through different concept features, attention is paid to convergent results from resonance scales and interviews. Eventually, the main insights will be incorporated into a product review presentation as design rationale. A preferred style of presenting the results is to show users' reactions to both prototype and competing products through video clips and transcribed quotes.

The data generated by RT are also useful for convincing stakeholders inside and outside the organisation about the qualities of the concept. This facilitates making decisions about continuing the concept development further by or with the client. It can help craft convincing concept statements, which can be used in market research. Later testing of buying intentions and product launch planning can benefit from extracting users' language from RT transcripts to describe the concepts as users explained them. This is particularly important for launching radically new products.

8 Case examples

We present two examples of RT that took place at Continuum between 2007 and 2009. The descriptions omit market research questions or usability evaluations sometimes included in the tests (e.g. likelihood of purchase, brand associations, convenience, expected price range, etc.). Products details are obscured for client confidentiality.

8.1 Premium personal healthcare

This project involved the design of packaging and display material for three series of products. The project targeted buyers of premium health and beauty care products with high household incomes. Three product series were screened for their specific groups of users with gender balanced 2:1 (women:men). The main test target was identification with the products. All prototypes were semi-functional, but they allowed using the product for real.

The test procedure included three levels of presentation: concepts as ideas, packaging display and experiencing the design by using it. Each level had different types of instruments to measure resonance. At the concept and packaging levels, the three new products were put on the same line with existing, competing products. Four resonance scales for novelty (*ordinary–intriguing*), monetary value (*costs more–costs less*) and personal relevance (*for me–for others* and *for everyone–not for anyone*) were used in both assessments. A quote from a male subject explaining their responses for a concept rated on the personal relevance scale: "*Looks like organic [product], peace, love, Grateful Dead, vegetarian; an anti-thesis of who I am… who is less worried about their own health and more worried about, you know, where these products are produced… Just not me*".

The resonance scales were complemented by questions about how much the users believed the product would cost and why they perceived its personal value as they did. In the packaging level test, users were also asked to match the packaging with the concept description to see if the packaging met the expectations created by the concept. In the experience test, users tested at least two of the three new concepts they preferred the most. After trying out the product as they would normally use it, they were asked whether the experience matched their expectations formed on the basis on the previous levels of presentation.

During the three iterations of RT, it was discovered that more technical cues and scientific language were needed to increase the credibility of the concepts, both in concept statements and in the prototypes. Resonance scales demonstrated that designs for premium brands did stand out visually, with some expected resistance from users. One of the three concepts was perceived by the users as representing a new, distinctive category in the domain. Testing suggested changes both to product and packaging features, for instance reducing granularity in paste consistency and removing a product container outline from the package. The interviews helped to pick out product properties that subjects had found distinctive for the new concepts. Overall, designers gained confidence in the written concept statements and could move forward with two concepts, one for each two out of three user segments.

8.2 Beverages for everyone and everywhere

This project involved creation of numerous new concepts for delivering beverages. The early design produced nearly 20 product prototypes for different user groups. The key attributes for design were freshness, sharing and ecological sustainability. These were subjected to RT for two broad groups: single adults and parents with children. This included cross-cultural testing in the USA and Italy. The majority of the artefacts were functional prototypes, although they did not contain the beverages.

The tests were customised for each user group because the experiential targets and concept for each group were different. This meant choosing different resonance scales and recruiting kids with a parent instead of individuals. For adults, the resonance scales focused on freshness, usability (handling, pouring, storage), disposal and environmental impact. All assessments were followed up by asking for a rationale for the rating to tease out why the users perceived the attributes as they did and how visual design implied certain characteristics. For kids, simplified resonance scales about the fit to the use context were used (for home, for school). Quote from a girl interviewed with her mother explaining her choices in the 'for school' scale: "*This one is the second best, because, like, when you do it, they [other kids at school] can't see it because they would think that it's silly*".

Findings from RT showed that users were very accustomed to certain use patterns, with clear, but not necessarily well-founded expectations about products and the use experience. This made it challenging to communicate the new type of designs to users heavily primed with existing offerings. However, forcing subjects to assess several new prototypes along with existing products helped them to articulate their implicit attitudes and assumptions about beverage containers beyond what could have been observed ethnographically or by testing the new offerings alone. It was also found that the Italian consumers were more receptive than Americans to a trusted seal than visual cues in conveying environmental footprint. It was also discovered which designs best communicated the concept of *freshness.* These design features were prioritised for further concept development.

9 Discussion and conclusions

In this paper, we have presented a concept evaluation method used in a global design company. It contributes to literature by combining three key features: (a) connecting design targets to evaluation, (b) joint evaluation of multiple concepts and (c) introducing resonance scales for measuring target implementation. We started off by a study of theoretical requirements for concept testing and demonstrating that no comparable, professionally-used methods are currently published in the product development research literature. We proceeded to review and analyse the high level goals for design, prominently desirability. We connected this analysis to the rationale of RT method, showing the need for an agile and qualitative feedback method in early design.

RT is presented here as an example practice of evaluating UX on concept level. For designers, RT is a qualitative method for design feedback and validation of user's perceptions. The value of RT comes from understanding why customers make certain rankings or categorise the concept as they do, which provides insights for concept revisions. It tells whether the design resonates; does it evoke the experiential responses and perceptions of functionality envisioned by designers. This helps to create designs that combine elements that resonate the strongest with the selected customer segment.

We have described why and how to use resonance testing in concept evaluation. The focus has been on how experiential and functional design targets are fulfilled. The method provides both design inspiration and decision making. It has been found useful also for generating user feedback for radically new products, something traditionally difficult. We have provided a solid description of the method, even though in our practice RT is used as a set of ingredients, from which we adapt the most suitable parts for every

situation. It has been argued that this is typical of industrial use of design and evaluation 'methods', as put by Woolrych et al. (2011, p.940): "*methods are loose incomplete collections of resources, which successful practitioners configure, adapt, and complement to match specific project circumstances*". We think the same way. Thus, it also important to describe practices with minimal abstraction as they are found in the industry to receive a realistic picture of practices influencing the innovation process (cf. Liikkanen et al., 2011)

9.1 Advantages of resonance testing

Resonance testing takes the stance that only by a multi-stage assessment, and using artefact comparison and stimulated interviews, we can quickly get to the bottom of the users' perception of new products. Although hardly any of the elements involved in RT is novel by itself, their unique combination provides an edge for RT. For instance, we believe that the presentation of familiar benchmark products is necessary to stimulate the right discourse. It allows people to discuss design features and characteristics they would otherwise be unable to consider, as they are unaware of the influence of visual design features (Veryzer and Hutchinson, 1998; Creusen and Schoormans, 2004).

In our opinion, other evaluation methods we presented earlier do not seem to provide similar design input as RT does. Among the widely utilised traditional marketing techniques, laddering suffers from the requirements of radical coding and later condensation (Grunert and Grunert, 1995). This risks losing contact between authentic chains of attributes, consequences and values when the interview data are synthesised. This relates to the flaws of applicability repeatedly observed among UX evaluation methods (Vermeeren et al., 2010). Another philosophical difference is that in RT we think that the attributes are relatively important. Designers make educated guesses about important design targets and RT includes laddering to understand why the artefacts match to targets differently. But RT offers more than laddering, which assumes considerable explicit prior knowledge, where as RT is driven by visual and tactile features.

The bipolar laddering variation is also of limited use. In the examples provided by Pifarré et al. (2009), it appears that functional, usability, image and 'emotional' problems get thus all mixed up together, disconnected from design. The authors of bipolar laddering also focus on individual and isolated products, notably those users have experience with, not new ones. The AXE method (Gegner and Runonen, 2012) has a complicated analysis scheme and does not promise clear design implications, reducing its desirability for design practitioners.

9.2 Future work

Resonance testing as a method has several points for development. The risks involved with RT include overgeneralisation with small samples and a confirmation bias to accept initial designs. Test user selection is also critical and not very well formalised. A study by Peng and Finn (2010) suggested that 'domain-specific innovativeness' discriminates among respondents on their capability to give feedback about concepts. This further calls for careful screening of users. On the other hand, the nature of the feedback may be different if it is provided by innovative, expert 'lead-users' (den Ouden et al., 2006;

Moore, 2002) instead of the majority. For the purposes of RT, responses from typical representatives of the user segment are important for final acceptance, because different user segments perceive and evaluate concepts differently.

The focus of RT on perception makes it dependent on visual and tangible features. Experiences from non-tangible, e.g. food products, have been somewhat disappointing, implying that the test is biased for visual features. This paper excluded any attempt of validating RT. This was an intentional omission as our goal was to present the ideas embedded in our practice. However, others may want to select some elements (Woolrych et al., 2011) of RT for detailed study. For instance, the use of resonance scales could be compared to related methods for feasibility and the quality of feedback. We see that the RT method, or practice, described here is too broad and interconnected with a specific design process to be validated as a whole.

9.3 Conclusion

We have introduced one way of performing user evaluation in concept design. Resonance testing is presented as a method that combines several important principles to help to gather meaningful feedback even for radically novel concepts, but it is not a fixed tool for every design situation and innovation challenge. It is a tool for making informed concept decisions and improving concepts quickly.

Acknowledgements

Brian Stonecipher, Harry West, Jake Childs and Chris Hosmer from Continuum all provided valuable input for this study. Author LAL was working as a visiting scholar at Stanford University H-STAR centre and the Department of Communications while preparing this paper and is indebted to the inspiring and hospitable environment within. Research project LUTUS by Technology Industries of Finland Centennial Foundation supported LAL.

References

al-Azzawi, A., Frohlich, D. and Wilson, M. (2008) 'User experience: a multiple sorting method based on personal construct theory', *UXEM Workshop at CHI'08*, Florence, Italy.

Artacho, M.A., Ballester, A. and Alcántara, E. (2009) 'Analysis of the impact of slight changes in product formal attributes on user's emotions and configuration of an emotional space for successful design', *Journal of Engineering Design*, Vol. 21, pp.693–705.

Benedek, J. and Miner, T. (2002) 'Measuring desirability: new methods for measuring desirability in the usability lab setting', *Usability Professionals Association Conference 2002*, Orlando, FL.

Brown, T. (2009) *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*, Harper Business, New York.

Chuang, M.C., Chang, C.C. and Hsu, S.H. (2001) 'Perceptual factors underlying user preferences toward product form of mobile phones', *International Journal of Industrial Ergonomics*, Vol. 27, pp.247–258.

Cooper, A. (2004) *The Inmates are Running the Asylum: Why High Tech Products Drive us Crazy and How to Restore the Sanity*, Pearson Higher Education, Indianapolis, IN.

Cooper, A., Reimann, R. and Dubberly, H. (2003) *About Face 2.0: The Essentials of Interaction Design*, John Wiley & Sons, Inc., New York.

Creusen, M.E.H. and Schoormans, J.P.L. (2004) 'The different roles of product appearance in consumer choice', *Journal of Product Innovation Management*, Vol. 22, pp.63–81.

den Ouden, E., Yuan, L., Sonnemans, P.J.M. and Brombacher, A.C. (2006) 'Quality and reliability problems from a consumer's perspective: an increasing problem overlooked by businesses?', *Quality and Reliability Engineering International*, Vol. 22, pp.821–838.

Dreyfuss, H. (1955) *Designing for People*, Simon and Schuster, New York.

Duke, C.R. (1994) 'Understanding customer abilities in product concept tests', *Journal of Product & Brand Management*, Vol. 3, pp.48–57.

Gegner, L. and Runonen, M. (2012) 'For what it is worth: anticipated experience evaluation', in Brassett, J., Hekkert, P., Ludden, G., Malpass, M. and Mcdonnell, J. (Eds): *International Design and Emotion Conference*, 11–14 September, Central Saint Martins University of the Arts London and the Design and Emotion Society, London, pp.1–12.

Grunert, K.G. and Grunert, S.C. (1995) 'Measuring subjective meaning structures by the laddering method: theoretical considerations and methodological problems', *International Journal of Research in Marketing*, Vol. 12, pp.209–225.

Hassenzahl, M. (2004) 'The interplay of beauty, goodness, and usability in interactive products', *Human-Computer Interaction*, Vol. 19, No. 4, pp.319–349.

Hassenzahl, M., Diefenbach, S. and Göritz, A. (2010) 'Needs, affect, and interactive products: facets of user experience', *Interacting with Computers*, Vol. 22, pp.353–362.

Hassenzahl, M. and Tractinsky, N. (2006) 'User experience: a research agenda', *Behaviour & Information Technology*, Vol. 25, pp.91–97.

Holtzblatt, K. (2011) 'What makes things cool? Intentional design for innovation', *Interactions*, Vol. 18, pp.40–47.

Hsu, S.H., Chuang, M.C. and Chang, C.C. (2000) 'A semantic differential study of designers' and users' product form perception', *International Journal of Industrial Ergonomics*, 25, pp.375–391.

Hummels, C. (2007) 'Searching for salient aspects of resonant interaction', *Knowledge, Technology & Policy*, Vol. 20, pp.19–29.

Iuso, B. (1975) 'Concept testing: an appropriate approach', *Journal of Marketing Research*, Vol. 12, pp.228–231.

Jordan, P.W. (2000) *Designing Pleasurable Products*, Taylor and Francis, New York.

Karapanos, E. and Martens, J-B. (2007) 'On the discrepancies between designers' and users' perceptions as antecedents of failures in motivating use', *IADIS International Conferences Interfaces and Human Computer Interaction 2007*, 3–8 July 2007, Lisbon, Portugal, pp.206–210.

Karapanos, E. and Martens, J.-B. (2008) 'The quantitative side of the repertory grid technique: some concerns', *UXEM workshop at CHI'08*, Florence, Italy.

Klemmer, S.R., Sinha, A.K., Chen, J., Landay, J.A., Aboobaker, N. and Wang, A. (2000) 'Suede: a wizard of Oz prototyping tool for speech user interfaces', *The 13th Annual ACM Symposium on User Interface Software and Technology (UIST'00)*, ACM Press, San Diego, CA.

Krippendorff, K. (2006) *The Semantic Turn: A New Foundation for Design*, CRC Press, Boca Raton, FL.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E. and Sinnelä, A. (2011) 'UX curve: a method for evaluating long-term user experience', *Interacting with Computers*, Vol. 23, pp.473–483.

Laakso, M. and Liikkanen, L.A. (2012) 'Dubious role of formal creativity techniques in professional design', *International Conference on Design Creativity (ICDC'2012)*, Glasgow, UK.

Lees, G. and Wright, M. (2004) 'The effect of concept formulation on concept test scores', *Journal of Product Innovation Management*, Vol. 21, pp.389–400.

Liikkanen, L.A., Laakso, M. and Björklund, T. (2011) 'Foundations for studying creative design practices', In Martens, J-B. and Markopoulus, P. (Eds): *Proceedings of the Second Conference on Creativity and Innovation in Design*, 19–21 October, ACM Press, New York, NY, pp.309–315.

Macdonald, A.S. (2001) 'Aesthetic intelligence: optimizing user-centred design', *Journal of Engineering Design*, Vol. 12, pp.37–45.

Moore, G.A. (2002) *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*, Harper Paperbacks.

Moore, W.L. (1982) 'Concept testing', *Journal of Business Research*, Vol. 10, pp.279–294.

Mueller, J.S., Melwani, S. and Goncalo, J.A. (2012) 'The bias against creativity: why people desire but reject creative ideas', *Psychological Science*, Vol. 23, pp.13–17.

Mullins, J.W. and Sutherland, D.J. (1998) 'New product development in rapidly changing markets: an exploratory study', *Journal of Product Innovation Management*, Vol. 15, pp.224–236.

Nagamachi, M. (2002) 'Kansei engineering in consumer product design', *Ergonomics in Design*, Vol. 10, pp.5–9.

Norman, D.A. (1988) *The Psychology of Everyday Things*, Basic Books, New York.

Norman, D.A. (2004) *Emotional Design: Why We Love (or Hate) Everyday Things*, Basic Books, New York.

Norman, D.A. (2010) 'The way I see it: technology first, needs last: the research-product gulf', *Interactions*, Vol. 17, pp.38–42.

Peng, L. and Finn, A. (2008) 'Concept testing: the state of contemporary practice', *Marketing Intelligence & Planning*, Vol. 26, pp.649–674.

Peng, L. and Finn, A. (2010) 'Whose crystal ball to choose? Individual difference in the generalizability of concept testing', *Journal of Product Innovation Management*, Vol. 27, pp.690–704.

Pifarré, M., Sorribas, X., Villegas, E., Fonseca, D. and García, O. (2009) 'BLA (bipolar laddering) applied to Youtube: performing postmodern psychology paradigms in user experience field', *Journal of Behavioral, Cognitive, Educational and Psychological Sciences*, 1, pp.76–82.

Pifarré, M. and Tomico, O. (2007) 'Bipolar laddering (BLA): a participatory subjective exploration method on user experience', *Conference on Designing for User eXperiences (DUX'07)*, 5–7 November, ACM Press, New York, NY, pp.2–13.

Raita, E. and Oulasvirta, A. (2011) 'Too good to be bad: favorable product expectations boost subjective usability ratings', *Interacting with Computers*, Vol. 23, pp.363–371.

Redström, J. (2006) 'Towards user design? On the shift from object to user as the subject of design', *Design Studies*, Vol. 27, pp.123–139.

Reynolds, T.J. and Gutman, J. (1988) 'Laddering theory, method, analysis, and interpretation', *Journal of Advertising Research*, Vol. 28, pp.11–31.

Roto, V., Ketola, P. and Huotari, S. (2008) 'User experience evaluation at Nokia', *UXEM Workshop at CHI'08*, Florence, Italy.

Roto, V., Law, E., Vermeeren, A. and Hoonout, J. (2011) *User Experience White Paper: Bringing Clarity to the Concept of User Experience*, Dagstuhl Seminar on Demarcating User Experience, 15–18 September.

Roto, V., Rantavuo, H. and Väänänen-Vainio-Mattila, K. (2009a) 'Evaluation user experience of early product concepts', *Designing Pleasurable Product and Interfaces (DPPI'09)*, 13–16 October, Compiegne University of Technology, Compiegne, France.

Roto, V., Väänänen-Vainio-Mattila, K., Law, E. and Vermeeren, A. (2009b) 'User experience evaluation methods in product development', *12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (UXEM'09)*, Springer-Verlag, Uppsala, Sweden.

Schoormans, J., Eenhuizen-van den Berge, M., van de Laar, G. and van den Berg-Weitzel, L. (2010) 'Design packages that communicate product attributes and brand values: an exploratory method', *The Design Journal*, 13, pp.31–47.

Smith, W.R. (1956) 'Product differentiation and market segmentation as alternative marketing strategies', *Journal of Marketing*, Vol. 21, pp.3–8.

Tan, F.B. and Hunter, M.G. (2002) 'The repertory grid technique: a method for the study of cognition in information systems', *Mis Quarterly*, Vol. 26, pp.39–57.

Trott, P. (2001) 'The role of market research in the development of discontinuous new products', *European Journal of Innovation Management*, Vol. 4, pp.117–126.

Väänänen-Vainio-Mattila, K., Roto, V. and Hassenzahl, M. (2008a) 'Now let's do it in practice: user experience evaluation methods in product development', *Extended Abstracts on Human Factors in Computing Systems (CHI'08 EA)*, ACM Press, Florence, Italy.

Väänänen-Vainio-Mattila, K., Roto, V. and Hassenzahl, M. (2008b) 'Towards practical user experience evaluation methods', *5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM)*, Reykjavik, Island.

Väätäjä, H., Olsson, T., Savioja, P. and Roto, V. (2012) 'UX Goals 2012 workshop – how to utilize user experience goals in design?', *NordiCHI2012*, 14 October 2012, Tampere University of Technology, Copenhagen, Denmark, p.48.

Veludo-De-Oliveira, T.M., Ikeda, A.A. and Campomar, M.C. (2006) 'Laddering in the practice of marketing research: barriers and solutions', *Qualitative Market Research: An International Journal*, Vol. 9, pp.297–306.

Vermeeren, A.P.O.S., Law, E.L-C., Roto, V., Obrist, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. (2010) 'User experience evaluation methods: current state and development needs', *The 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI'10)*, ACM, Reykjavik, Iceland.

Veryzer, R.W.J. and Hutchinson, J.W. (1998) 'The influence of unity and prototypicality on aesthetic responses to new product designs', *Journal of Consumer Research*, Vol. 24, pp.374–385.

Wherry, J.S. (2006) *Simulated Test Marketing: Its Evolution and Current State in the Industry*, Master of Business Administration Master's, Massachusetts Institute of Technology.

Williams, D., Kelly, G. and Anderson, L. (2004) 'MSN 9: new user-centered desirability methods produce compelling visual design', *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, Vienna, Austria.

Wind, Y. (1973) 'A new procedure for concept evaluation', *Journal of Marketing*, Vol. 37, pp.2–11.

Woolrych, A., Hornbaek, K., Frökjaer, E. and Cockton, G. 2011) 'Ingredients and meals rather than recipes: a proposal for research that does not treat usability evaluation methods as indivisible wholes', *International Journal of Human-Computer Interaction*, Vol. 27, pp.940–970.

Note

1 All UX evaluation methods. Available online at: http://www.allaboutux.org/all-methods (accessed on 14 March 2014).