
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Kujala, Sari; Walsh, Tanja; Nurkka, Piia; Crisan, Marian

Sentence completion for understanding users and evaluating user experience

Published in:
Interacting with Computers

DOI:
[10.1093/iwc/iwt036](https://doi.org/10.1093/iwc/iwt036)

Published: 01/01/2014

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Kujala, S., Walsh, T., Nurkka, P., & Crisan, M. (2014). Sentence completion for understanding users and evaluating user experience. *Interacting with Computers*, 26(3), 238-255. <https://doi.org/10.1093/iwc/iwt036>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

The original place of publication in *Interacting with Computers*, 2014, 26, 3, pp. 238-255:

<http://iwc.oxfordjournals.org/content/26/3/238.full.pdf+html>

Sentence completion for understanding users and evaluating user experience

Sari Kujala^a, Tanja Walsh^b, Piia Nurkka^b, Marian Crisan^b

^aAalto University, School of Arts, Design and Architecture, PO Box 31000, FI-00076 Aalto, Finland, firstname.secondname@aalto.fi

^bTampere University of Technology, Human-Centered Technology, PO Box 589, FI-33101 Tampere, Finland, fistname.secondname@tut.fi

Abstract: Projective techniques are used in psychology and consumer research to provide information about individuals' motivations, thoughts, and feelings. This paper reviews the use of projective techniques in marketing research and user experience research and discusses their potential role in understanding users, their needs and values, and evaluating user experience in practical product development contexts. A projective technique called sentence completion is evaluated through three case studies. Sentence completion produces qualitative data about users' views in a structured form. The results are less time-consuming to analyze than interview results. Compared to quantitative methods such as AttrakDiff, the results are more time consuming to analyze, but more information is retrieved on negative feelings. The results show that sentence completion is useful in understanding users' perceptions and that the technique can be used to complement other methods. Sentence completion can also be used online to reach wider user groups.

Keywords: User experience evaluation, projective techniques, sentence completion, user values

1. Introduction

As Hassenzahl and Tractinsky (2006) describe, user experience (UX) goes beyond the task-oriented approach of traditional HCI and focuses on hedonic aspects of use such as fun and pleasure. The hedonic aspects of user experience are those that satisfy universal human needs but do not necessarily have any utility value. As Vyas and van der Veer (2006) point out, products may be liked not only because they offer functional features but also because of other irrational and subjective reasons. They give the example of how an Apple iPod is not just a portable music player; it also represents users' social status.

User experience is a vague concept and there is little common agreement on its full nature and scope (Law et al., 2009). The ISO standard 9241-110 (2010) defines user experience broadly as the users' perceptions and responses in regard to their interaction with a system or product. Based on Law's et al. (2009) survey of 275 user experience researchers and practitioners, user experience is

agreed to be dynamic, context-dependent, and subjective. Many researchers also highlight the importance of emotions (Forlizzi and Battarbee, 2004; Hassenzahl and Tractinsky, 2006; Isomursu et al., 2007; Mahlke, 2005) and the holistic and phenomenological nature of user experience (McCarthy and Wright, 2004; Swallow et al., 2005).

It is known that people do not just passively undergo emotional experiences, but they actively interpret the meaning of these experiences and construct memories of them (Holland and Kensinger, 2010). Accordingly, user experience is frequently seen as constructive (Battarbee, 2003; Sanders and Dandavate, 1999; Vyas and van der Veer, 2006; Wright et al., 2003) and sense making (McCarthy and Wright, 2004). For example, Vyas and van der Veer (2006) state that users are not concerned with products as such, but with the values and meanings products bring to their lives.

The evaluation of user experience is challenging as users may find it difficult to express their experiences if directly asked to. As Vermeeren et al. (2010) point out user experience is subjective, so objective usability measures such as task execution time or the number of errors are not sufficient for user experience evaluation. The solution is to develop not only quantitative questionnaire measures such as Finstad's (2010) usability metric for user experience, but also practical methods to determine what is relevant from the user point of view and how users interpret and reflect their experiences and create meanings with a product.

There are a wide variety of user experience evaluation methods (Vermeeren et al., 2010;argas-Avila and Hornbæk, 2011). One frequently used method (Bargas-Avila and Hornbæk, 2011), is the AttrakDiff semantic differential questionnaire developed by Hassenzahl (2004), which provides quantitative information about user experience. However, as user experience is personal, holistic and complex, it is difficult to determine whether users find the predefined properties essential and meaningful to them. In addition, because the predefined measures may not reveal all the aspects of user experience, many researchers like open, qualitative evaluation methods (Vermeeren et al., 2010). Furthermore, the quantitative results do not explain the reasons behind the ratings – the very reasons, which would provide developers direct input on how to improve their designs.

Furthermore, although qualitative methods such as interviews can be used, certain dimensions of experience cannot directly be asked (Springett, 2009). For example, it is known that for social desirability reasons, some users do not mention status or prestige issues in self-reports (Richins, 1994). Springett (2009) points out that direct questioning tends to lead to general statements of user attitudes and reactions without efficiently unpacking the reasons behind them. To deeply understand user experience, it is necessary to find out users' interpretations and views and to circumvent the problems of direct questioning created by social barriers and researchers' presuppositions.

Several non-verbal methods have been developed to measure users' emotional responses to products, for example SAM (Bradley and Lang, 1994) and PrEmo (Desmet, 2003). However, while measuring users' emotional responses is useful when comparing two different designs, it does not provide information about the reasons for the emotional responses and therefore may be less applicable for product development (DiSalvo et al., 2004).

The openness and experimental format of cultural probes may be seen as particularly suited for non-task-focused parts of user experience (Boehner et al., 2007; Swallow et al., 2005) and some researchers have developed new user experience evaluation techniques inspired by cultural probes. For example, Swallow et al. (2005) developed user-directed evaluation of user experience called "Do Something..." challenges. The participants are asked to select five items from a list of emotional adjectives and carry out activities with the product that are related to the adjectives. The

idea is that an open-ended exercise promotes the use of the product in inventive ways and encourages reflection on the feelings that the product use evokes. The technique seems to provide interesting feedback from users with the five usage scenarios selected, but it is intended to be supplemented with anticipation and reflection interviews.

In psychology, projective techniques have been used to bypass or circumvent the conscious defenses of respondents to gain unconscious information from them. Projective techniques typically present respondents with an ambiguous stimulus, such as an inkblot, and ask them to describe what they see or to generate a response (Lilienfeld et al., 2000). The rationale here is that as the stimulus itself has little meaning, the respondents are interpreting it based on their own views.

In product development and user experience evaluation, projective techniques are not well-known and are rarely utilized, although they seem to provide a promising approach for understanding user motivations and interpretations of their experiences with products. The purpose of this study is to improve understanding of projective techniques and their possible role in understanding users and evaluating user experience. Accordingly, firstly the experiences of using these techniques in marketing and user experience research are reviewed; and secondly, the results of applying a projective technique - sentence completion - in understanding user needs and values and evaluating user experience are presented.

1.1 Projective techniques in consumer and marketing research

Projective techniques have also been adapted for consumer and marketing research to study user motivation, thoughts, feelings, beliefs, attitudes and experiences (Chang, 2001; Donoghue, 2000; Steinman, 2009; Will et al. 1996). The types of techniques used are slightly different from those used in psychology. Consumer researchers do not employ ambiguous stimulus material that may produce a wide variety of answers but more practically direct respondents to produce material related to a particular research topic. The techniques used include association (connecting the research object with words, images, or thoughts), story construction, bubble drawings (completing speech bubbles in a cartoon), and completion (finishing sentences or stories) (Donoghue, 2000). Even "expressive" methods such as role-playing are mentioned (Will et al, 1996; Donoghue, 2000).

Hoyer and MacInnis (2007, p. 60) give an example of a completion technique and describe a study in which cigarette smokers were asked why they smoked. Most of the smokers said they enjoyed it and believed that smoking in moderation was fine. However, when they were given incomplete sentences like "People who never smoke are _____", Respondents filled in the blanks with words like *happier* and *wiser*. And when given sentences like "Teenagers who smoke are _____", respondents answered with words like *crazy* and *foolish*. Hoyer and MacInnis (2007) conclude that these smokers were clearly more concerned about smoking than their explicit answers indicated.

Zinkhan et al. (1999) give an example of a more product-oriented use of a projective technique. They used a story completion-based projective technique to learn about consumer motivations for constructing personal Web pages. The respondents were exposed to a scenario in which an imaginary person purportedly displayed affect related to the development of a Web page. Then the 90 respondents were asked to write a brief story or commentary about the likely motives of the imaginary Web page creator. Using the technique Zinkhan et al. (1999) were able to identify several classes of motivations and interestingly, less than 20% of the responses were related to rational or utilitarian reasons for creating personal pages. For example, creating Web pages was seen to satisfy the need to construct and convey creators' identities and the need for power. Thus, as the authors conclude, the technique seemed to help in identifying hidden or socially sensitive motivations although it can't be known how general the motivations are.

Boddy (2005) and Steinman (2009) reviewed the research related to the validity and reliability of projective techniques in consumer research. Haire's (1950) shopping list study is the first published study on projective techniques. He asked 100 women to write a brief description of a woman's personality and characteristics based on two versions of a shopping list. The lists were identical, except that one included Nescafé instant coffee and another traditional coffee. The shoppers of traditional coffee were described in a more positive manner whilst the Nescafé shoppers were described as lazy and inefficient household planners and schedulers. Haire's (1950) shopping list study was performed in the United States, but the study has been replicated in other locations such as Norway and France and the original findings have remained consistent across international replications (see Steinman, 2009 for review).

Both Boddy (2005) and Steinman (2009) state that Haire's (1950) study and its numerous replications indicate that projective techniques are a useful approach to better understand consumer feelings. Boddy (2005) also finds many positive reports by market research practitioners of the usefulness of projective techniques. However, both Boddy (2005) and Steinman (2009) agree that research into the reliability and validity of using projective techniques is lacking. Steinman (2009) argues that despite the popularity of projective techniques, the small sample sizes and monetary and time commitments associated with them may have limited publications in academic journals.

In summary, varied projective techniques are used in consumer research and the results are found useful. Several advantages in using projective techniques have been identified, including the amount and richness of the information collected. Compared to direct questioning using attitude scales, projective techniques yield a wider range of responses providing understanding of a consumer's thoughts and feelings, experiences, and motives (Donoghue, 2000; Doherty and Nelson, 2010). Indirect questioning also reduces the social desirability bias that results from respondents wishing to avoid embarrassment and project a favorable image to others (Fisher, 1993; Doherty and Nelson, 2010; Will et al., 1996). In addition, projective techniques are often inspiring to users as using them is fun and engaging for both participants and researcher alike (Doherty and Nelson, 2010; Will et al., 1996) and often stimulates discussion (Will et al, 1996). The primary disadvantage of projective techniques found is the degree of difficulty and subjectivity related to the interpretation of the complex data (Doherty and Nelson, 2010; Will et al, 1996; Steinman, 2009). In addition, the reliability of measures is difficult to establish (Donoghue, 2000).

The experiences of projective techniques in consumer research are positive and thus, it seems likely that the same techniques could be applied also for evaluating user experience.

1.2 Projective techniques in user experience research

As mentioned, projective techniques are relatively unknown in HCI and user experience research. Design-oriented cultural probes, initially developed by Bill Gaver (1999) and others, have some similarities to projective techniques. Cultural probes are a set of packages of, for example, maps, postcards, and other materials combined with tasks designed specifically for the research setting and which aim to provoke responses from the users (e.g. Gaver et al., 1999). Cultural probes are widely used and numerous variations have been developed (Boehner et al., 2007). The original cultural probes were not designed to provide data about users, but rather to spark design inspiration along with personal communication (Gaver et al., 1999); as such they often focus on emotional aspects of interaction design (Boehner et al., 2007). That approach was similar to the constructive type of projective techniques used in consumer research as users were given ready-made materials and asked to construct something describing their experiences. Projective techniques in consumer research in contrast, aim to understand consumer motivations and experiences.

Cultural probes are often adapted to provide information that will give clearer guidance to the design process as a type of user requirements gathering or collecting feedback on a particular application (Boehner et al., 2007). Many employ probes as a means of supporting interviews, or else interviews are used to support the interpretation of probe results. These kinds of information-oriented probe variants tend to present more clear-cut and focused questions to obtain better data. Boehner et al. (2007) argue that these focused studies work against the original idea of breaking preconceptions and creating personal and sensitive communication between designers and users. However, there are several needs for both cultural probes and projective techniques. There is a need to spark inspiration in the early concept design. Furthermore, product development companies need more systematic techniques for gathering user feedback related to the experiential aspects of use that can be generalized to one or several target user groups.

A study by McDonagh et al. (2002) is one of the rare attempts to utilize projective techniques in design. They adapted the projective technique of product personality profiling (PPP) from market research to design in order to elicit user perceptions and emotional responses to products. Participants were asked to imagine a product as a person with a particular personality, and provide information regarding its character and lifestyle. McDonagh et al. found that the technique provides an insight into how the user perceives to be the target consumer and it helps to reveal social value systems and emotional responses to products. The limitations of the technique they mention are that interpretation of the results may be complex and may lead to over-using stereotypes.

In summary, projective techniques are not currently widely used in user experience research except in the form of cultural probes to obtain inspiration for design and to establish a conversation with users. However, looking at the advantages found through their use in consumer research, projective techniques could potentially be useful for understanding user experience. Since user experience is multidimensional and complex and the relevance of the dimensions can vary from one product and user to another, it is difficult to design good scales. In some situations therefore, it may be more useful to gather inspirational data, stimulate discussion and get a deeper understanding about what is important to users and how they interpret their experiences with products. Projective techniques could be useful for understanding user needs and values as well as gathering user feedback on existing prototypes and products.

1.3 Sentence completion

In this study, we explore the possibilities of a semi-structured projective technique called sentence completion in understanding user needs and values and evaluating user experience. This technique is well established in both psychology and marketing: for example, Soley and Smith (2008) reviewed some studies showing that the results of sentence completion correlate with real life behavior by the respondents.

Sentence completion is a combination of a projective technique and a questionnaire, whereby respondents are provided with beginnings of sentences (called sentence stems) that they then complete in ways that are meaningful to them (Soley and Smith, 2008, p. 132). Sentence completion can be used to assess variety of constructs, including motivations and attitudes (Soley and Smith, 2008, p. 131). By providing only the beginning of the sentence, a researcher gives the topic, but respondents have the freedom to respond to it as they wish.

Users are instructed to answer quickly and without thinking too much as there are no correct answers. Sentence completion does not give ready answer categories that could reveal the researcher's point of view about the topic but users are invited to interpret the sentence stimulus from their own perspective. Soley and Smith (2008, p. 144) pointed out that the sentence

completion tests appear to be more useful across cultures than bipolar scales, because they are less likely to be culturally biased. Users are using their own words and they are instructed to describe their first reactions to, and associations with, the given topics.

Sentence completion has also some practical benefits, especially compared to interviews. The answers are given in a written form which helps the analysis work and it is useful in industrial contexts where relatively fast and easy to apply methods are needed (Väänänen-Vainio-Mattila et al., 2008) and in international settings where the need for remote methods that can be used online has been identified (Monahan et al., 2008).

2. Research method

The sentence completion technique was applied adopting the multiple-case research design strategy (Yin, 1994; Dubé and Paré, 2003). Multiple case studies were performed to understand the influence of variability in context and to gain more generic research results (Yin, 1994).

The goal was to adapt sentence completion as a means to understand user needs and values and evaluate user experience and test the suitability of the technique in practical cases. As shown in Table 1, the cases represent diverse applications covering various dimensions of user experience, all of which enhances the generalizability of the results (Eisenhard, 1989). Apart from sentence completion, additional methods were used to validate the results and to allow comparisons of the usefulness of different methods. The preliminary results of sentence completion use in Case Study 1 and 2 have been published in Kujala and Nurkka (2009) and Walsh et al. (2010) respectively. We have also applied sentence completion for identifying needs and values related to computer usage in a parallel case study published in Nurkka et al. (2009).

Table 1. Summary of case studies.

Case	Application	Practical problem	Methods	Number of Resp.	Target of evaluation
I	Game	Designing a game that motivates children to exercise and is acceptable to parents.	Sentence completion, interviewing	10	Understanding users: needs and values
II	Smartphone	Evaluating UX of a Smartphone in a late phase of product development with a cross-cultural sample.	Remote sentence completion web questionnaire	97	Evaluating UX
III	Mobile phone camera	Evaluating the usability and UX of the camera in two different mobile phones.	Sentence completion, AttrakDiff, usability testing, interviews	20	Evaluating UX

2.1 Developing the sentence completion technique for understanding users and evaluating user experience

The development of the sentence completion technique started from scratch, as it had not been used before for understanding users or evaluating user experience. The aim was to gather practical qualitative data for product development purposes. Rather than considering the sentence completion technique to be used alone, it was seen as a supplementary technique to support other methods. The technique could be used for identifying users' views about what is relevant in their experience of a certain product or to stimulate discussion between users and designers.

The first sentences were developed relating to user needs and values associated with user experience (Hassenzahl et al., 2010; Kujala and Väänänen-Vainio-Mattila, 2009). As Hassenzahl et al. (2010) showed, the fulfillment of universal psychological needs is the major source of positive experience with products; accordingly they described universal psychological needs as competence, relatedness, popularity, stimulation, meaning, security, or autonomy. The same psychological needs have also been called user values in the context of using products (Kujala and Väänänen-Vainio-Mattila, 2009), human needs (Oulasvirta and Blom, 2007) and concerns (Desmet and Hekkert, 2007). Identifying these user needs and values is important both in the early phases of the design process and later when evaluating the user experience.

When sentence completion is used for evaluating user experience, the beginnings of the sentences used should focus on respondents' personal experiences and perceptions of the product or service under evaluation. As there were no existing sentences, we created some related to the user experience dimensions identified in the literature (Jordan, 2000; Hassenzahl and Tractinsky, 2006; Desmet and Hekkert, 2007). Although the set of sentences may not represent all dimensions of user experience, our focus was initially adapting sentence completion for gathering user information and feedback.

2.2 Ensuring validity of sentence completion

Although we were developing a sentence completion technique for product development purposes and not as a strict psychometric measurement tool, we followed the recommendations of Lilienfeld et al. (2000) so as to develop as valid a technique as possible. First, Lilienfeld et al. (2000) recommend that the projective technique should have a number of items designed to assess the same construct as this means measurement error can typically be averaged out by aggregating across multiple items. Thus, we created several sentence stems (beginnings of sentences) representing each user experience dimension. As the beginning of the sentences also guide respondents, we started the evaluation with general sentences like 'Using the product is ___'. The general sentences are open enough that they do not lead respondents to answer according to any preconceived expectations. Later in the questionnaires, more specific topics were probed such as 'The appearance of the product is ___'. In this way the questionnaire included multiple items for potentially gathering information related to varied user experience dimensions.

Second, Lilienfeld et al. (2000) recommend that the projective technique should consist of ambiguous stimuli that are especially relevant to the construct being assessed. In the case of sentence completion, the stimulus was specifically selected to be related to user values and user experience. Sentences are a more direct stimulus compared to the Rorschach inkblots used in psychology, so it is quite straight-forward to develop relevant stimulus material. In addition, as the concepts of user needs, user values and user experience are still abstract and not clearly defined, the first more open sentence stems ensured that the most important issues from a user point of view were also identified. These first sentence stems were more general to give respondents more freedom to describe their experiences in their own words, but the later sentence stems were more

focused on certain dimensions of user experience. In each case, the sentence stimulus material was also piloted to make sure that it was understood as intended and based on the feedback and results the sentences that inspired users to produce responses were selected for the actual study.

Third, as Lilienfeld et al. (2000) recommend, we used an iterative and self-correcting approach to developing sentence completion whereby, the developer begins with a tentative formulation of the construct to be assessed and then progressively revises these constructs and the stimuli assessing them on the basis of new data. The cross-case research supported the iterative development of sentence completion. The sentence completion material was developed and piloted for each case study. Feedback was gathered and the sentences were iteratively developed each time.

3. Cases: understanding users and evaluating user experience

The sentence completion technique was applied and developed through several case studies.

3.1 Case Study 1

Case Study 1 tested the sentence completion technique in an industrial partner's development case. The goal of the case study was to identify needs, values and attitudes that parents hold regarding their children's spare time and activities. Because the industrial partner's goal was to develop a game that would encourage children to exercise more, the values of parents were particularly interesting as not only is raising children value loaded from a parent's point of view, but also the parent's values may be expected to affect product acceptance and user experience. In addition to sentence completion, the parents and children were also interviewed. Here, however, we discuss only the results of the parents' interview and sentence completion results.

Although interviewing and sentence completion are different approaches, we wanted to compare the results in order to see how well sentence completion provides new information. The strength of interviews is in openness of the approach; semi-structured interviews can provide information that the interviewer did not expect. On the other hand, it was not known how well the open sentence completion task would work. Although the interviews could certainly provide rich information about additional aspects such as context of use, we focused on user needs and values as this was the main target of the study.

3.1.1. Methods

Ten families were recruited for the study through a circle of acquaintances of the researchers, mailing lists and through a newspaper ad. There were three single-mother families and seven nuclear families with one to four children whose ages ranged from 5 to 14 years. The age of the parents ranged from 31 to 45 years and they represented various professions. Participants were rewarded with movie tickets.

One parent per family completed 24 incomplete sentences (Appendix A). The sentences were constructed based on the value categories identified by Kujala and Väänänen-Vainio-Mattila (2009). The focus was on respondents' needs and values relating to their children exercising. All the interviews followed the same structure: first there were general questions about the family and about the parent's own free-time (e.g. what the family is like, what they do together, what kind of interests and hobbies the parent has), then there were general wellness and exercise related questions (e.g. does any family members exercise, what motivates her/him to exercise, do they have any sports gear or devices). The last half of the questions were child related questions probing a wide range of topics – from the child's personality, general hobbies, and exercise habits to aspects that might motivate /demotivate the child to exercise.

At the beginning of the interview session, one parent (8 mothers and 2 fathers) was asked to fill in the sentence completion questionnaire and a background information form. After that they were interviewed. With four families both parents were present at the interviews. The interviews lasted from one to one and half hours.

All the interviews were audio-recorded and transcribed. The results were analyzed in three steps. First, user values were identified and enumerated based on the categories formed by Kujala and Väänänen-Vainio-Mattila (2009). To ensure the reliability of the results, two researchers independently analyzed them. The researchers identified user values and counted them, both in sentence completion and the related interviews. Next, because the user value categories identified by Kujala and Väänänen-Vainio-Mattila (2009) partly overlap, the original value categories were further modified to be more explicit and to better describe the identified values related to children exercising: 1) Emotional and hedonistic values, 2) Social values, 3) Self-actualization and achievement, 4) Good behavior, 5) Safety, and 6) Well-being. Two researchers then independently categorized the identified values using the new coding categories and one outside coder who had not participated in the study categorized 50% of the data. An interrater reliability analysis using the Kappa statistic was performed to determine the consistence of the identified value categories among raters. Third, the value categories identified in the interviews and sentence completion were compared to find any similarities and differences between the methods.

3.1.2. Results

The respondents completed 223 sentences (92.9%) out of the 240 original sentence stems. There were only 76 (34%) one or two word responses, most were full sentences. The length of the responses varied depending on the sentence stem type. For example, the sentence stem ‘My child is most interested in’ provided only one or two word responses, whereas the sentence stem ‘My child’s best experience was...’ provided long responses such as ‘staying over night in a tent in their own garden which was exciting enough and safe’. Table 2 shows more examples of the parents’ responses to the sentence completion.

Table 2. Examples of the parents’ responses.

Sentence	Response
The most important thing to me is...	... the well-being of me and my family.
My child is most interested in...	... computer games.
My children exercise...	... irregularly, but willingly when they have company.
It is important in my child's physical activities that...	... she enjoys it and it is good for her health.
My child receives positive attention in physical activities if...	... he participates or succeeds e.g. meets a goal.
The emotion my children’s exercising arouses...	... is that I should encourage them more to regularly exercise.
My child’s best experience was...	... down-hill skiing as there were friends.
Regarding my child's physical activities I want to know...	... the company he has and how he is behaving.

The parents stressed emotional, hedonic and social values. They wanted their child to have fun and to have positive experiences. For example, they responded that it is important that the child's physical activities include: ‘joy’, ‘joy and positive experience of the self’, ‘that she enjoys it, has fun’ and ‘that she likes it’. They also wanted the family or child and friends to do things together. For example, eight parents replied that friends and/or family encourage the child to exercise. In

addition, the answers revealed that self-actualization and achievement were important; for example, the parents replied that their child gets positive attention in physical activities if ‘she succeeds, e.g. makes a goal’, ‘he tries and keeps up with others’, ‘he succeeds’, and ‘he does his best’.

The sentence completion results were compared to the interview results. The numbers of identified user values were very consistent between the two raters, the Pearson correlation coefficient was .93 ($p=.000$) for interviews and .96 ($p=.000$) for sentence completion. Slightly more user values were identified with the sentence completion technique ($M=12.1$, $SD=3.5$) than with interviewing ($M=10.4$, $SD=4.6$). However, the difference was not statistically significant with ten participants ($p=.091$, Wilcoxon test).

In content analysis, we were interested to discover which value categories each of the methods could reveal. The interrater reliability for the two raters was found to be almost perfect for the interviews (Kappa = .822) and substantial for sentence completion (Kappa = .710) using Landis and Koch’s (1977) guideline for interpretation. The interrater reliability for the three raters and 50% of the data was satisfactory for the interviews (Kappa = .706) and very good for sentence completion (Kappa = .828).

The content analysis showed that the interviews and sentence completion provided similar results. Both methods revealed all the value categories, but when the replies of individual participants were analyzed, altogether 38 out of the 54 (71%) identified categories were same in the interviews and sentence completion. Some value categories were identified more often in only one of the methods (See Table 3). The parents mentioned practical safety and good behavior more often during interviews than in sentence completion. In the interview, the parents may have had more time to consider different topics and both safety and behavior are socially accepted topics that parents may be supposed to speak about. On the other hand, in sentence completion, the parents focused slightly more on well-being and emotional and hedonic values. In all except one interview, the participants mentioned emotional and hedonic values, however these values were mentioned almost twice as often in sentence completion (42 times) than in the interviews (25 times).

Table 3. The value categories identified by interviewing and sentence completion (1=Emotional and hedonistic values, 2=Social values, 3=Self-actualization and achievement, 4=Good behavior, 5=Safety, 6=Well-being).

ID	Interview	Sentence completion
1	1, 2, 3, 4, 5, 6	1, 2, 4, 5, 6
2	1, 2, 3, 5, 6	1, 2, 3, 6
3	1, 2, 3, 6	1, 2, 3, 4, 6
4	2, 4, 6	1, 2, 3, 4, 6
5	1, 2, 3, 4, 6	1, 2, 3, 4, 6
6	1, 2, 3, 4, 5	1, 2, 3, 4
7	1, 2, 3, 4, 5	1, 2, 3, 6
8	1, 2, 3, 4, 5, 6	1, 2, 3, 6
9	1, 2, 3, 4, 5, 6	1, 2, 3, 6
10	1, 2, 3, 4	1, 3, 6

Sentence completion seemed to help one participant to express a sensitive issue better than in the interview as she mentioned how children tease each other on a general level: “Teasing relates to team sports, self-esteem is affected. It probably relates to physical education in schools.” In sentence completion she responded that her child is prevented from exercising by ‘friends? – weather’ and that the child’s exercise is embarrassing if ‘he stays beside, does not participate although he can’ and the emotion my child’s exercising arouses is ‘I am nervous for him, I am upset when he does not participate’.

Interviewing required more time than gathering data with sentence completion did. Each interview lasted from one to one-and-a-half hours, while the sentence completion took from five to fifteen minutes. Furthermore, the analysis of the interviews was more demanding as the long audio recordings needed to be transcribed and there was more material to analyze. For example, the outside researcher spent an average of 20 minutes analyzing an interview, but an average of 4 minutes analyzing sentence completion replies.

On the other hand, formulating the sentence stimulus material for the first time took more time, piloting, and developing iteratively the sentence stems took more time than preparing the interview questions. Moreover, the interviews did provide more detailed information about user needs and context of use. For example, the interviews at the families’ homes told us more about the environment where the children play and the children’s favorite activities and sports equipment.

3.1.3 Conclusions

This small exploratory case study showed that sentence completion was effective in identifying user needs and values related to children’s exercise. The results were a good starting point for designing a game concept that is attractive to both parents and children. The comparison of these results with the interview results showed that sentence completion provided similar information about user

needs and values in a fraction of the time and that the analysis was also more straightforward. However, interviews provided more information about user behavior and context of use that was not probed for by the sentences. Formulating the sentence stimulus material and piloting it also took more time than preparing interview questions.

3.2 Case Study 2

In Case Study 2 we tested sentence completion in our industrial partner's product development case. The goal was to explore how applicable a remote online sentence completion questionnaire would be in a cross-cultural sample evaluating user experience of a Smartphone. Our industrial partner, whose products are sold worldwide, was interested to gain input to designers about different dimensions of user experience and to develop their remote user experience evaluation toolset.

Globalization and the search for experiential aspects of technology products and services have increased the demand for cross-cultural user feedback we wished to obtain here. Fast and reliable user data collection enables designers to gain understanding of users from different locations and contexts. This makes remote methods, especially online questionnaires, attractive for global user data collection as they are practical, cost-effective and wide scale (Väänänen-Vainio-Mattila et al., 2008). Monahan et al. (2008) reported that no common practice currently exists on how user experience can be studied remotely in international contexts and therefore the development of suitable remote methods is needed. As mentioned earlier the sentence completion tests appear to be more useful across cultures than bipolar scales, because they are less likely to be culturally biased (Soley and Smith, 2008, p. 144). Therefore, it was found attractive for gathering cross-cultural user feedback.

3.2.1 Methods

The standard online questionnaire procedure of the industrial partner was used to perform the study. One hundred and thirty of our industrial partner's employees from 10 different countries, including China, Denmark, Finland, Germany, India, Japan, Singapore, the UK, the USA and the United Arab Emirates, were invited to answer to an online sentence completion questionnaire. The participants were volunteers and they were not rewarded. The response rate was 75% (97 out of 130). The Smartphone evaluated was a prototype in a late phase of product development and which the respondents had used for 5 months. Most users came from the USA (26%) or China (17%). Respondents included engineers, developers and business people, the majority of them were men and most were between 25-39 years old. The survey was in English and the users were able to answer fluently in English. The survey was open in a web-tool for 7 days.

The remote online questionnaire comprised of 14 sentences to be completed (See Appendix B). The sentence stems were developed to evaluate various dimensions of user experience presented in the literature (Jordan, 2000; Hassenzahl, 2004; Hassenzahl and Tractinsky, 2006; Desmet and Hekkert, 2007): Usability (Sentences 9, 10, 14), utility (Sentence 2), visual/haptic/acoustic aesthetics (Sentence 5 and 6), identification (Sentence 12), and socio-pleasure (Sentences 11 and 13). Mostly the sentence stems were very open focusing on experiential aspects 'how the product feels' and the respondents could describe the varied user experience dimensions with their own words. In some sentences the user experience dimensions e.g. aesthetics were directly probed with the sentence stems. Identification was probed by Sentence 12 describing the typical owner of the product as a product provides identification by expressing one's self (Hassenzahl, 2004) and it is known that the respondents find it easier to express this kind of status issues when they are thinking other people than themselves (Richins, 1994).

Users were asked to complete the sentences in the following way: “Please complete the sentences so that they describe how you feel. Respond quickly without thinking too long. You can leave a sentence without an answer if you feel it is not relevant to your situation”. At the end of the survey, users were asked to give open feedback to the product development team.

An affinity diagram (Holzblatt et al., 2005) was used in the analysis of the data from the completed sentences in order to categorize the individual responses for each of the sentence completion task (see an example in Figure 1). The higher level topics that emerged were labeled using the voice of the respondent. A user could also have had more than one answer item in one sentence, e.g. when asked to complete “Using my Smartphone is...” one participant answered “enjoyable, but sometimes frustrating”, where *enjoyable* is one answer item and *frustrating* another answer item.

3.2.2 Results

The respondents completed 1234 sentences (90.9%), forming 171 answer categories (See Appendix B). Table 4 presents examples of the sentence responses and Figure 1 gives an example of the resulting answer categories. According to our industrial partner, during previous product evaluations similar kinds of user groups had not been keen on giving answers to open questions. Therefore, our industrial partner was surprised how sentence completion tasks were successful in eliciting such a large amount of free-text answers. The large number of completed answers indicates that the sentence completion method inspired respondents to communicate their product experiences as reported feedback and as a result we were able to get information about users’ personal feelings and thoughts from different user experience dimensions.

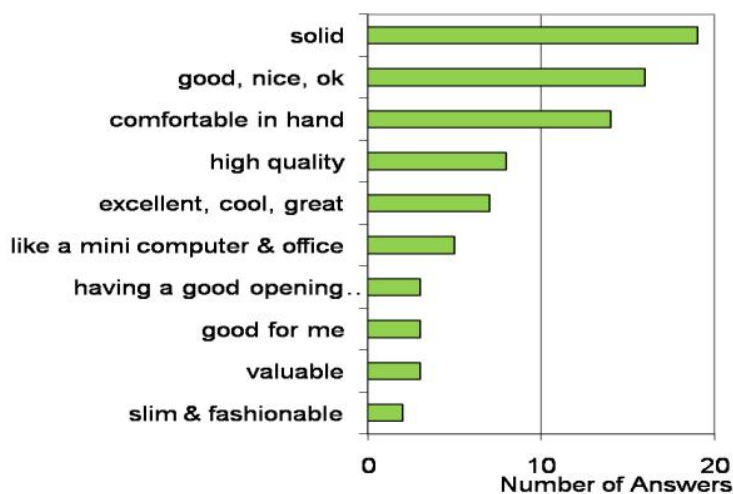


Figure 1. The categories formed from sentence completion results with sentence number 6 (My smartphone feels...). *Smartphone* refers here to the name of the Smartphone under evaluation.

Sentence completion yielded data for product user experience analysis and ideas for design. The summary of the Smartphone user experience (see Table 5) were presented to and discussed with the product developers. The findings were mostly related to usability, but usability and user experience intertwine especially in positive issues. For example, the users report that the phone feels professional and other people want to buy it when it is cheaper (an example in Table 4). This is related to identification and socio-pleasure, the phone supports users professional identity and it is presentable in front of other people. Also sleek and elegant refers to that. The phone is not just an efficient tool, but it feels positive and powerful.

Table 5. A summary of the Smartphone user experience evaluation

Positive		Negative
Professional & confident		Buggy
Form & size feels comfortable in hand		Slow UI SW
Keep in touch & communicate		HW heats up
High quality opening mechanism	Summary of Smartphone User Experience	Not as expected
Rich & diverse functions		Complicated to use
Supporting my daily life		Battery life too short
Browsing & e-mailing		Touch responsiveness
Sleek & elegant		Music & Video & Games
Touch screen		Not good for quick/rough everyday use
Useful		Feeling stressed & frustrated

Interestingly, according to our industrial partner, some answer categories found in the data were new to them and yielded new points of view about the product's user experience. From product development point of view, feedback and problems related to technical issues, utility and usability can often be found rather fast with a smaller testing team, but intangible user experience is more challenging to evaluate. Thus, the answers related to visual/haptic/acoustic aesthetics (Sentences 5 and 6), identification (Sentence 12), and socio-pleasure (Sentences 11 and 13) were bringing new insight. For example, these sentences provided information on how and in which ways the device was meaningful for the users and supported their daily life. The responses also revealed how the product reflected the feeling of professionalism and how it made them feel confident. The answers categorized in "Not as expected" gave insight into issues related not only to utility or usability but more also to what qualities in the device made them feel that it was not as they expected. This new information was important in product development itself but it can also be used when designing future user experience evaluation surveys.

Case Study 2 demonstrated that a remote online sentence completion questionnaire is a relatively fast and easy way of gathering international user data – providing the users have access to a computer and the internet and are able to understand and write in English if the questionnaire can't be translated.

3.2.3. Conclusions

In Case Study 2 we used a remote online sentence completion questionnaire to collect a large amount of qualitative data about the different dimensions of user experience of a Smartphone for product development purposes. Case Study 2 demonstrated that sentence completion can be effectively performed online to quickly gather useful international data. The study results helped the industrial partner to identify essential user experience issues from user point of view and they provided feedback about different dimensions of product user experience. Sentence completion can be applied in online questionnaires, which is an important factor in international studies or when large user groups need to be reached.

3.3 Case Study 3

Case Study 3 evaluated usability and user experience of two mobile phone cameras and compared the results of usability testing, an adapted version of the AttrakDiff questionnaire (Hassenzahl et al., 2003; Hassenzahl, 2004) and sentence completion. The AttrakDiff was selected as it is one of the most frequently used user experience evaluation methods (Bargas-Avila and Hornbæk, 2011). The goal was compare the usefulness of sentence completion and the AttrakDiff in evaluating user experience.

3.3.1 Methods

Although the 20 participants lived in Finland, they included exchange students or immigrants from Africa, Anglo-America, Asia, Eastern Europe and Latin America. There were five participants from each part of the world. Fourteen were men and the mean age of the participants was 25.0 years. None of the participants were previously familiar with the phones. Participants were rewarded with movie tickets.

Ten of the participants evaluated the camera on the basic Nokia 3720 classic mobile phone and ten participants evaluated the camera on the Nokia N86 smartphone.

The sentence completion material was based on the questions used in Case Studies 1 and 2, but the number of sentences was reduced to 16 in order to avoid participant fatigue as three methods were used during the same test session. The questionnaire included 16 sentences to be completed. Only 15 of the 28 AttrakDiff items were selected for the questionnaire so that the evaluation session would not be too long for the participants (see Figure 2). As the cameras are inside the mobile phone, the items related to appearance were not valid and could not be used. In addition, as the internal consistencies of the AttrakDiff's three scales are high, Cronbach's alphas were .85, .95 and .90 (Hassenzahl, 2004), the shortened version was expected to provide approximately same results as the long AttrakDiff. The removed items were simple, direct, manageable, integrating, classy, valuable, inclusive, presentable, aesthetic, inviting, sympathetic, and motivating.

Based on a pilot test, the wording of the three of the 15 items was modified to make them clearer to cross-cultural users. 'Brings me closer' was changed to 'supporting social life', 'human' was changed to 'human-centered', 'easy-challenging' was changed to 'easy to use-difficult to use'. In addition, two slightly new word pairs were added: 'useful-useless' and 'fun-boring'. The scale used was from one to six instead of from one to seven so as to avoid respondents selecting excessively neutral options as seen in the study of Lee et al. (2002).

The evaluation began with a standard usability test in which users performed eight common phone camera-based tasks, including taking photographs and transferring them to a computer. The tasks were always presented in the same order. After the usability test, the user experience of the phone cameras was evaluated by sentence completion and the AttrakDiff questionnaire (Hassenzahl, 2004). Finally, the participants were interviewed.

Two researchers used an affinity diagram method to analyze the data from the completed sentences (Holzblatt et al., 2005) in order to categorize the individual responses.

3.3.2 Results

Using the phone cameras appeared to be slightly complicated in the usability test with users encountering an average of 7.4 usability problems. For example, eight users had difficulties finding the camera and two users could not find it without help.

In spite of the problems with basic functions, the users rated N86 very positively with the adapted AttrakDiff (see Figure 2). Nokia 3720 was rated to be rather typical and amateurish but easy to use and practical. The users considered the phone cameras to be easy to use. Comparing the sentence completion answers with the AttrakDiff ratings, the respondents gave a similar rather positive view. For example, the African users of N86 responded: ‘When I use this phone camera I feel...very happy everything going on nice’ and ‘...cool and can use it anywhere’. The responses of the sentence ‘Using this phone camera is...’ were categorized into four kinds: Easy (6 responses), Slightly disappointing (5 responses), OK (4 responses), Nice/Comfortable (4 responses).

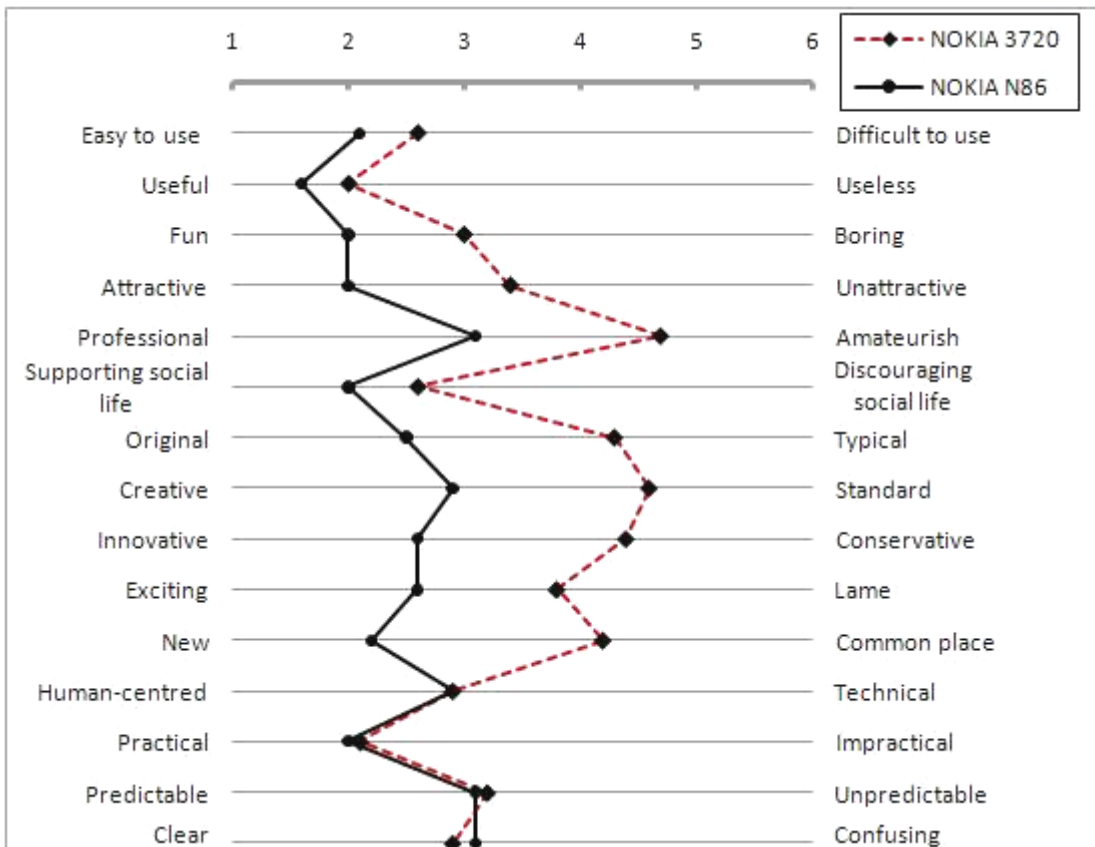


Figure 2. Means of the adapted AttrakDiff questionnaire responses.

However, there were also some negative user experiences. One reason is that sentence completion directly probed for possible problems and all except one of the users mentioned a problem such as a difficult to use function (twelve users) or low quality (seven users). Fifteen of the problem responses gave specific feedback to the designers, such as ‘the zoom function is hidden’, ‘you can’t take 2 pics fastly’, ‘the arrows are not clear’, ‘taking two pictures in short succession is clumsy’ and ‘the camera button feels wrong’. In addition, the sentence ‘This phone camera is not suitable for...’ elicited the downsides of the phones such as ‘taking pictures from far’ and ‘transfer of computer’, and ‘taking photos of landscapes (the quality of little details is poor)’.

In addition to the two problem related sentences, six Nokia 3720 and two N86 users mentioned something clearly negative in other responses. For example, they replied that ‘using it is sort of difficult because of multifunction’, ‘using it is difficult in the beginning, because it doesn’t have the shortcut to the camera’, ‘people think that the pictures will not be good enough’, ‘people think their images are darker’ and ‘people think that I take it with an old camera’. One of the Latin Americans gave the most negative replies saying that ‘using it is not good’, ‘the problem is quality’ and ‘when using the phone camera he feels frustrated’ and ‘other people think that he can’t buy a good phone’.

The same person also had difficulties finding the camera in the usability test while he rated the phone camera to be extremely ease to use, very practical and rather clear with AttrakDiff.

Particularly striking is that two Africans rated the N86 phone camera as being very easy to use and practical with AttrakDiff (both gave 1), although they faced more usability problems than other users. Furthermore, although their sentence completion answers were overall very positive, they did provide some explanations: for example, the African who had the highest number of usability problems said that using the mobile phone for the first time is not easy, but at the same time when using it he feels cool and can use it anywhere. Thus, it looks like that because the mobile phone was so 'cool' they forgave the difficulties with it or considered that the reasons for the difficulties lay with themselves.

As the Nokia 3720 was a basic model, five of the users described its identification and typical owners in slightly negative terms: 'middle aged', 'who don't use it', 'older... my parents' age', 'people who don't usually take photos with a phone camera'. On the hand, people from Eastern Europe seemed to appreciate the basic model more, describing the typical owner to be: 'average person, no gadget owner' and 'a normal person who prefer to have a phone for basic necessities (not to show off and not to have the phone only for calls and SMS)'.

3.3.3 Conclusions

Usability testing, sentence completion and AttrakDiff provided slightly different feedback about the two mobile phones. As expected, usability testing was clearly best for providing details of the usability problems and their reasons. Some information was also received about users' likes and dislikes: for example, many complained that the phone camera took too long to take a picture.

Both AttrakDiff and sentence completion provided more feedback on the hedonic dimensions of user experience compared to usability testing. AttrakDiff provided very easy to analyze results in a structured manner. Users seemed to overrate the ease of use with AttrakDiff. This may be because it may not be nice to admit to having difficulties with a product, or sometimes users did not notice usability problems; some even claimed the problem lay with themselves and not the product.

Compared to AttrakDiff, sentence completion was better for gathering negative feedback and reasons for the negative feelings. In an extreme case, the respondent described how 'using the phone feels frustrating' in the sentence completion and still he rated the phone to be very easy to use in the AttrakDiff questionnaire. It should be noted however, that because the results of this Case Study were qualitative, the end analysis demanded more effort.

4. Discussion and Conclusions

In this article we have discussed the potential role of projective techniques in understanding users; what is important to them, their needs and values and how they interpret their experiences with products. Projective techniques are based on the idea that individuals' characteristics, needs, and life experiences influence their interpretation of ambiguous stimuli and that by using projective techniques the problems of direct questioning can therefore be circumvented. Use of these techniques in marketing research suggested that they could potentially be useful for providing the information about individuals' motivations, thoughts and feelings that are relevant in examining user experience.

This study presents our first steps in developing a sentence completion technique for identifying user needs and values and evaluating user experience, taking the recommendations of Lilienfeld's et al. (2000) as a foundation. Our goal was to develop a practical technique for gathering qualitative

user feedback for product development purposes rather than to develop a strict measurement tool. Sentence completion was selected as it appears practical to use in product development contexts. As users are completing given sentences, this technique is not as ambiguous as other projective techniques, the responses can be focused on the preferred topics, responses are received in a structured written form and the technique can also be used online to reach more representative user groups.

The three case studies (summarized in Table 6) demonstrated the use of sentence completion for different purposes in product development. The technique was used for identifying user needs and values in the concept design phase (Case Study I), gathering information about user experience after trial use (Case Study II) and evaluating user experience in conjunction with usability testing (Case Study III). Sentence completion was shown to be useful for these different purposes as well as identifying different dimensions of user experience (User values in Case I, different user experience dimensions in Cases II and III).

Table 6. Summary of case study results.

Case	Results
I	Compared to interviewing, sentence completion provided similar information about user needs and values. Formulating the sentence stimulus material for the first time and piloting it took more time compared to interviewing, but otherwise sentence completion required less time for performing and analyzing the results. The results provided important information for designing a game concept.
II	The remote online sentence completion questionnaire was quick and useful for gathering a large volume of qualitative and quantitative international data about different dimensions of product user experience.
III	Sentence completion provided similar results to the adapted AttrakDiff questionnaire when used for evaluating user experience, but was better for gathering negative feedback and reasons for evaluations and emotions (although analyzing the results was more demanding. Both methods provided more feedback on the hedonic dimensions of user experience than usability testing did. As expected usability testing was best for identifying detailed usability problems.

The results show that sentence completion is useful in gathering information about user needs and values and experiential aspects of use and that the technique can be used to complement other methods. A particular strength of the technique is that users use their own words, and thus compared to multiple-choice or semantic differential questions, they can much more openly express their values and how a product feels and what kinds of associations it has. However, compared to open-ended techniques, sentence completion is semi-structured, the ready sentence stems guide users to specific themes.

As Case Study I showed, sentence completion is effective in gathering focused data, but it requires less time to perform and to analyze the results than interviewing does. Formulating the sentence stimulus materials required effort for the first time and piloting the materials took some time. We hope that the ready sentence stems provide a good starting point for creating materials for new studies. Many of the sentence stems can be easily reformulated for a new context by just changing the product name or the topic in the sentence.

Case Study 2 demonstrated that a good representative set of users can be easily reached online. Although field studies and usability tests provide rich and high-quality data, in practice only a limited number of users can participate and users may not be willing to raise sensitive issues face-to-face. In product development, there is utility in gathering cost-efficient international user feedback, and since it is not often possible to travel to many locations, online methods are necessary to complement basic methods (Luedemann & Muller, 2010).

As mentioned online questionnaires are considered to be cost-effective, because data collection is faster and demands less of a work force (eg. Väänänen-Vainio-Mattila et al., 2008). Firstly, data is ready for analysis immediately after delivery (Fan & Yan, 2010). Secondly, the samples can be bigger and the increased sample size does not make much difference to the total cost of the study as it would with traditional methods (Benfield & Szlemko, 2006). Thirdly, it is easier to reach respondents worldwide (Evans & Mathur, 2005). Internet reduces the time and distance between people, and makes communication more efficient (Ekman & Litton, 2007). In addition, online questionnaires may be more convenient to respondent as they can answer when suitable for them. The disadvantage is that the quality of the research may be threatened by errors in the coverage: not everybody has access or uses Internet as Internet penetration is not evenly distributed across segments of the population. Furthermore, when studying user experience, non-response should be considered as a possible threat to the validity of the study. For example, if only respondents having positive experiences respond, the understanding of the UX remains one-dimensional.

The results of Case Study 3 suggest that in contrast to the adapted AttrakDiff questionnaire, sentence completion provides more qualitative negative feedback, information that is useful for improving user experience. For example, the users gave sometimes information about the specific function causing trouble like “The zoom function is hidden”. This kind of feedback from users provides developers more direct input on how to improve user experience. However, as the sentence completion results were qualitative, their analysis demanded more effort.

Sentence completion can provide both the qualitative and quantitative data needed in practical product development in a rather structured way and as such readymade sentences would be useful and relatively easy to apply in practice. However, as user experience goals differ from product type to another, it is challenging to provide a list of sentences and a scale for analyzing the results that would be applicable to all situations. In addition, understanding of the dimensions of user experience is still evolving and new sentences should be created as new knowledge is gained. Furthermore, users use varied wordings for describing their experiences and automatic semantic analysis tools need to be developed to support analysis work in product development contexts.

The qualitative approach requires extra effort and interpretation in analysis as the researcher must categorize and summarize the results based on meaning. Therefore, although the analysis is relatively straightforward, it demands some interpretation and work, and can be a threat to validity. In order to avoid this threat and to keep the users’ voices alive, we used the users’ wordings as in-vivo codes in analysis as far as possible. Thus, when the responses were categorized, the categories were named based on the most commonly used wordings. Using the open general sentences in the beginning may improve the quality of the responses. Users can describe those experiences that are most important to them and it can be ensured that users are not being lead too much. For example, in Case Study 1 the sentences did not probe anything related to wellbeing but the general sentence “It is important in my child's physical activities that...” revealed that many of the parents considered wellbeing very important in their child’s exercising.

The use of written words helps the analysis work, but the verbal approach has also some limitations. Individuals differ in their ability to express themselves verbally. As user experience is holistic in nature, people may also find it difficult to reflect on it. Furthermore, because users have to evaluate their user experience based on their memory of the momentary experiences (Hassenzahl and Ullrich, 2007), they may not remember small details. Users may also be unaware of small details that may have had impact on overall user experience. For example, in Case Study 3, some users did not realize that they had a usability problem with the camera (they thought the camera did not have a zoom, but actually they simply could not find it). Case Study 1 also showed that interviewing is more efficient for understanding practical needs (e.g. safety) and context of use. Thus, the sentence completion technique would not be ideal for performing detailed analysis of context of use or ongoing experiences with product properties. Instead, the technique may be more suitable for getting a quick overall understanding of how users interpret their experiences and what is uppermost in their minds when they are evaluating the product.

The quality and richness of the responses produced by sentence completion varied. When users described a product they owned such as a mobile phone and they had used the product for a long period, they produced richer answers. Because they had a personal and long-lasting relationship with the product, the users had much more to say about their experiences in Case Study II than after the short usability test in Case Study III. It should also be noted that in Case Studies II and III, where the respondents were from several countries but the sentence completion was performed in English that those respondents using English as their native language produced more vivid and longer responses.

One of the challenges of sentence completion is to prioritize users' responses and find out which of the responses are most relevant from the user point of view (e.g. mentioned user values). One effective approach would be to use sentence completion to identify the relevant issues and a questionnaire to get more quantitative data about the importance of the issues among a wider set of users. We also found out that the order of sentences had an effect on the responses. The respondents kind of continued the story created by the sentences. Thus, it would be better if the positive or negative versions of the sentence stem are not placed adjacent to one another as the respondent may overly attempt to stay consistent with her replies.

Although sentence completion is a projective technique, it is not a direct way of revealing users' subconscious or innermost thoughts and feelings. If users wish, they are still able to select their responses and filter out issues that they do not want to reveal. As mentioned, the weakness of the technique is also that it requires participants to verbalize their thoughts. However, the sentence completion technique has some aspects that support users in openly expressing their experiences:

- The technique is associative in nature and it supports remembering experiences related to the product. Human memory is essentially associative, linking together related words and objects and when retrieved one item can cue the recall of related items (Mayes et al., 2007). Thus, the stimulus words in sentence stems activate the memories related to the topic and to personal experiences.
- The sentence completion tasks are designed to include very open sentence stems such as 'Using my Smartphone is...' and 'It is important in my child's physical activities that...' that give the respondent the possibility to reflect their experiences in whatever way they wish.
- A user experience sentence which probed for the typical user of the product helps people to express the status related issues when they are thinking about other people than themselves (Richins, 1994). In fact, one typical technique employed in market research

when measuring the image of the product is to ask a participant to describe the characteristics of the typical user (Allen et al., 2008).

- Using the tool asynchronously as in Case Study 2, the users can respond anonymously without the researcher being present.
- Based on our observations and the good response rates completing sentences seems to be a fun and motivating task for most of the respondents.

The first steps to ensuring the validity of the sentence completion technique and the resulting data were performed using the recommendations of Lilienfeld et al. (2000). The sentence stimulus material was iteratively developed and piloted. We learned how to formulate good quality sentence stems. For example, only one-word stems or long ones did not stimulate responses well enough. Formulating new sentence stems must be done carefully: the sentence stems need to be fluent to read and include a specific stimulus word e.g. the product name, but the stems need to be open enough that they fit all situations. In addition, the stimulus materials need to be piloted. The case studies provide a basic set of stimulus materials that can be used in future studies.

We did not develop an analysis key or categorization, so the sentence completion technique developed here is not a diagnostic quantitative tool. The sentences can be used directly as informational and inspirational data as or they can be categorized based on frequency to provide quantitative data. The frequency data is important as companies often appreciate quantitative data that can be used to evaluate the applicability of the results among wider numbers of users.

The case studies were designed to examine the strengths and limitations of the sentence completion technique for gathering feedback from users and as such they do not prove the approach's definitive validity. Moreover, before systematically examining the validity and reliability of the approach, we first needed to see whether the technique can be used in the user experience context and develop it further for this purpose. After the three case studies, the technique appears to be a promising practical tool to be further tested, but as our theoretical understanding of user experience is still evolving, the stimulus sentences and their coverage should be checked and developed further based on the new theory.

By comparing different methods, it was possible to see the differences in the quality and coverage of the data obtained. For example, the results show that the sentence completion technique seems to provide more qualitative negative feedback compared to the AttrakDiff. Since designers need concrete examples of users' experiences, sentence completion therefore seems to be useful. However, the challenge in determining the validity of the information is that there is no baseline for checking the correctness of the data. In addition, although the sentence completion technique is very open to the interpretations of the respondents, it is not known if it reveals all the relevant information. The technique may be better suited for identifying what is important and critical to respondents than providing information comprehensively from all topics. A quantitative comparison of different measures could show the relationship of sentence completion results to satisfaction, but as sentence completion provides mostly qualitative data this would not be a straightforward comparison.

Systematic research is therefore needed to ensure the validity and reliability of the sentence completion approach. Lilienfeld et al. (2000) provide some guidelines for future research. First, there is a need to test construct validity. It means for example that the sentence completion test should be able to differentiate between two fundamentally different user groups. Even more important is that the technique has a predictive value in real world settings. There may not be any need for a diagnostic tool as such, but the goal is to provide feedback for design purposes. The

predictive value would then mean that the technique provides relevant data and designers can really improve user experience based on the user feedback obtained.

In addition to validity issues, we can borrow from the psychometric literature and examine whether the sentence completion technique satisfies such criteria as 1) reliability e.g. test-retest reliability, interrater reliability, and internal consistency, 2) incremental validity e.g. the extent to which an instrument contributes information above and beyond other information (Lilienfeld et al., 2000). However, as the goal is to develop a practical evaluation method for design purposes, we can question to what extent these criteria are applicable. Test-retest reliability may not be relevant, as user experience is evolving over time, so it could be rather artificial to try to measure it. A certain degree of interrater reliability though, is desirable for user experience evaluation methods, so that the results do not depend solely upon the person interpreting them. Internal consistency may also be slightly pointless as the sentence completion tool aim to cover different aspects of user experience. On the other hand, incremental quality has pragmatic importance as real product development contexts have serious time and resource limitations. Taking these factors into account then, a good evaluation method should provide useful information above and beyond more easily collected data. This means that a method providing less information can still be useful in practice if it demands less resources (cf. Vermeeren et al., 2010).

In conclusion, this paper reports the first experiences of using sentence completion to identify user needs and values and to evaluate user experience. The case studies reported provide examples how to use sentence completion. Sentence completion produces qualitative data about users' views, but it is more structured than other projective techniques or cultural probes. The results are less time-consuming to analyze than interview results are. Compared to quantitative methods such as AttrakDiff, the results are more time-consuming to analyze, but more data is produced regarding the reasons for the negative feelings (Case 3). Sentence completion therefore, could complement other methods in practical product development, as it can provide qualitative data in a structured way and it can be used online to easily reach both a wide and distant audience of users.

ACKNOWLEDGEMENTS

We thank Hannu Soronen, Liinu Helkiö and Iiro Viitanen for their assistance in carrying out the studies. This research was supported by the Finnish Funding Agency for Technology and Innovation (Tekes) (VALU 40364/06, SUXES 40079/09) and the Finnish Doctoral Program in User-Centered Information Technology (UCIT).

6. References

- Allen, M. W., Gupta, R., & Monnier, A. (2008). The Interactive Effect of Cultural Symbols and Human Values on Taste Evaluation. *Journal of Consumer Research*, 35, 2, 294-308.
- Bargas-Avila, J.A. and Hornbæk, K., 2011. Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In *CHI 2011 Proceedings*, ACM, pp. 2689-2697.
- Battarbee, K., 2003. Defining co-experience. *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pp. 109–113.
- Benfield, J., A. and Szlemko, W., J. Internet-Based Data Collection: promises and Realities. *Journal of Research Practice* 2, 2, Article D1 (2006).
- Boddy, C., 2005. Projective techniques in market research: valueless subjectivity or insightful reality. *International Journal of Market Research*, 47, 3, 239–254.

- Boehner, K., Vertesi, J., Sengers, P., Dourish, P., 2007. How HCI interprets the probes. In CHI Proceedings, ACM, pp. 1077-1086.
- Bradley, M.M. and Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 1, pp. 45-59.
- Chang, J. E., 2001. The revival of projective techniques: Past, present, and future perspectives. *Advances in Consumer Research*, 28, 1, 253-254.
- Desmet, P.M.A., 2003. Measuring emotion; development and application of an instrument to measure emotional responses to products. In M.A. Blythe, A.F. Monk, K. Overbeeke, and P.C. Wright (Eds.), *Funology: from usability to enjoyment*. Dodrecht: Kluwer Academic Publishers, pp. 111-123.
- Desmet, P.M.A. and Hekkert, P., 2007. Framework of product experience. *International Journal of Design*, 1, 1, 57-66.
- DiSalvo, C., Hanington, B., Forlizzi, J., 2004. An accessible framework of emotional experiences for new product conception. In McDonagh, D., Hekkert, P., van Erp, J. and Gyi, D. (Eds.) *Design and Emotion, The Experience of Everyday Things*. London: Taylor & Francis, pp. 251-255.
- Doherty, S. and Nelson, R., 2010. Using projective techniques to tap into consumers' feelings, perceptions and attitudes . . . getting an honest opinion. *International Journal of Consumer Studies*, 34, 400-404.
- Donoghue, S., 2000. Projective techniques in consumer research. *Journal of Family Ecology and Consumer Sciences*, 28, 47-53.
- Dubé , L. and Paré, G., 2003. Rigor in information systems positivist case research: current practices, trends, and recommendations. *MIS Quarterly*, 27, 597–636.
- Eisenhardt, K.M., 1989. Building theories from case study research. *Academy of Management Review*, 14, 532–550.
- Ekman, L. and Litton, J-E., 2007. New times, new trends; e-epidemiology. *Eur. J. Epidemiology* 22, 285-292.
- Evans, J. and Mathur, A., 2005. The value of online surveys. *Internet Research* 15, 2, 195-219.
- Fan, W. and Yan, Z., 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26, 2, 132-139.
- Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22, 5, 323–327.
- Fisher, R.J., 1993. Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 20, 303-315.
- Forlizzi, J. and Battarbee, K., 2004. Understanding experience in interactive systems. In *Proceedings of DIS Conference*, ACM Press, pp. 261-268.
- Gaver, B., Dunne, T. and Pacenti, E., 1999. Cultural probes. *Interactions*, 6, 1, 21-29.

- Haire, M., 1950. Projective techniques in marketing research. *Journal of Marketing*, 14, 649-656.
- Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19, 329-349.
- Hassenzahl, M., Burmester, M., Koller, F., 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. [AttrakDiff: A questionnaire for the measurement of perceived hedonic and pragmatic quality]. In: Ziegler, J., Szwillus, G. (eds.) *Mensch & Computer 2003: Interaktion in Bewegung*, Teubner, Stuttgart, Germany, pp. 187–196.
- Hassenzahl, M., Diefenbach, S. and Göritz, A., 2010. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*, 22, 353-362.
- Hassenzahl, M., Tractinsky, N., 2006. User experience - a research agenda. *Behaviour & Information Technology*, 25, 2, 91-97.
- Hassenzahl, M. and Ullrich, D., 2007. To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19, 4, 429–437.
- Holland A. C. and Kensinger, E. A., 2010. Emotion and autobiographical memory. *Physics of life reviews*, 7, 1, 88–131.
- Holtzblatt, K., Burns Wendell, J. and Wood, S., 2005. *Rapid Contextual Design. A How-To Guide to Key Techniques for User-Centered Design*. Elsevier, San Francisco, USA, 159-179.
- Hoyer, W.D., MacInnis, D. J., 2007. *Consumer Behavior*. Houghton Mifflin Company, New York, USA.
- ISO DIS 9241-210 2010. Ergonomics of human system interaction – Part 210: Human-centred design for interactive systems. Switzerland: International Standardization Organization (ISO).
- Isomursu, M. Tähti, M. Väinämö, S. and Kuutti, K., 2007. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human-Computer Studies*, 65, 4, 404–418.
- Jordan, P.W. (2000). *Designing Pleasurable Products. An Introduction to the new human factors*. London, UK: Taylor & Francis.
- Kujala, S. and Nurkka, P. 2009. Identifying user values for an activating game for children. In *Proceedings of the MindTrek Conference: Everyday Life in the Ubiquitous Era*, New York: ACM Press, pp. 98-105.
- Kujala, S. and Väänänen-Vainio-Mattila, K., 2009. Value of Information Systems and Products: Understanding the Users' Perspective and Values. *JITTA: Journal of Information Technology Theory & Application*, 9, 4, 23-39.
- Landis, J.R., Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, 1, 159-174.
- Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J., 2009. Understanding, scoping and defining user experience: a survey approach. *Proceedings of the 27th international conference on Human factors in computing systems*, pp. 719–728.

- Lee, I. Jones, P., Minneyman, Y. Zhang, X., 2002. Cultural Differences in Responses to a Lickert. *Research in Nursing & Health*, 2002, 25, pp. 295-306.
- Lilienfeld, S.O., Wood, J.M. and Garb, H.N., 2000. The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27-66.
- Luedemann, I. and Muller, T., 2010. Online Surveys. In Schumacher, R. (Ed.) *The Handbook of Global User Research*. Morgan Kaufmann, Burlington, USA.
- Mahlke, S., 2005. Understanding users' experience of interaction. In *Proceedings of the 2005 annual conference on European association of cognitive ergonomics*, pp. 251–254.
- Mayes, A., Montaldi, D., & Migo, E., 2007. Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11, 3, 126–135.
- McDonagh, D., Bruseberg, A. and Haslam, C., 2002. Visual product evaluation: exploring users' emotional relationships with products. *Applied Ergonomics*, 33, 231-240.
- McCarthy, J., and Wright, P. 2004. *Technology as Experience*. Cambridge, MA, USA: MIT Press.
- Monahan, K., Lähteenmäki, M., McDonald, S. and Cockton, G., 2008. An Investigation into the Use of Field Methods in the Design and Evaluation of Interactive Systems. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, interaction - Volume 1*. British Computer Society, Swinton, UK, 99-108.
- Nurkka, P., Kujala, S., & Kempainen, K. (2009). Capturing users' perceptions of valuable experience and meaning. *Journal of Engineering Design*, 20, 5, 449-465.
- Oulasvirta, A., Blom, J., 2007. Motivations in personalization behaviour. *Interacting with Computers*, 20, pp. 1-16.
- Richins, M., 1994. Valuing things: The public and private meanings of possessions. *Journal of Consumer Research*, December, 21, 504-521.
- Sanders, E.B.-N., & Dandavate, U., 1999. Design for experiencing: New tools. In *Proceedings of the First International Conference on Design and Emotion*, Delft, pp. 87–92.
- Soley, L.C. and Smith, A.L., 2008. *Projective techniques for social science and business research*. The Southshore Press: U.S.A. Milwaukee.
- Springett, M., 2009. Evaluating cause and effect in user experience. *Digital Creativity*, 20, 3, 197-204.
- Steinman, R.B., 2009. Projective techniques in consumer research. *International Bulletin of Business Administration*, 5, 37-45.
- Swallow, D., Blythe, M., & Wright, P., 2005. Grounding experience: relating theory and method to evaluate the user experience of smartphones. *Proceedings of the 2005 annual conference on European association of cognitive ergonomics*, pp. 91–98.
- Vermeeren, A.P.O.S., Law, W.L.-C., Roto, V. Obrist, M., Hoonhaut, J., Väänänen-Vainio-Mattila, K., 2010. User experience evaluation methods: Current state and development needs. In *Proceedings of the NordiCHI conference*, pp. 521-530.

Vyas, D. and van der Veer, G. C., 2006. Rich evaluations of entertainment experience: bridging the interpretational gap. Proceedings of the 13th European conference on Cognitive ergonomics: trust and control in complex socio-technical systems, pp. 137–144.

Väänänen-Vainio-Mattila, K., Roto, V., and Hassenzahl, M., 2008. Now Let's Do It in Practice: User Experience Evaluation Methods in Product Development. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 – 10, 2008). CHI '08. ACM, New York, NY, 3961-3964.

Walsh, T., Nurkka, P. and Kujala, S. 2010. Exploring Cross-Cultural UX Evaluation with the Remote Online Sentence Completion Method. In Proceedings of the Ninth International Workshop on Internationalization of Products and Systems IWIPS, pp. 11-22.

Will, V., Eadie, D., MacAskill, S., 1996. Projective and enabling techniques explored. *Marketing Intelligence & Planning*, 14, 6, 38-43.

Wright, P., McCarthy, J., MeekiSon, L., 2003. Making Sense of Experience. In Blythe, M. (Ed) *Funology– From usability to enjoyment*. The Netherlands: Kluwer Academic Publishers, pp. 43-53.

Yin, R.K., 1994. *Case study research, design and methods*. 2nd ed. Beverly Hills, CA: Sage Publications.

Zinkhan, G.M, Conchar, M., Gupta, A. and Geissler, G., 1999. Motivations Underlying the Creation of Personal Web Pages: An Exploratory Study. *Advances in Consumer Research* 26, pp. 69-74.

Appendix A. Sentence completion form used in Case 1.

Please, complete the sentences below so that they describe you and your family. There is no wrong replies, respond rather quickly without thinking too long. You can leave a sentence without an answer if you feel that it is not suitable for your situation. By a child we mean here your school-aged children. If you have several school-aged children, please think about one of them.

The most important thing to me is _____

Our family could be described _____

My child is most interested in _____

My child's best experience was _____ because _____

Computer or electronic games _____

I give positive attention to my child if _____

My child exercises _____

My child receives positive attention in physical activities if _____

It is important in my child's physical activities that _____

From my child's point of view, the most important thing related to physical activities is _____

I want that by exercising my child _____

My child's exercise is particularly successful, if _____

The problem with my child's exercise is _____

The emotion my child's exercising arouses _____

My child's exercise is not pleasant if _____

My child's exercise is embarrassing if _____

My child's exercise is unpleasant if _____

My child is encouraged to exercise by _____

My child is prevented from exercising by _____

My own role in my child's exercise is _____

Competition in my child's exercise _____

I wish that when exercising my child looks like _____

Regarding my child's physical activities I would like to know _____

In the future, I wish my child _____

Thank you!

Appendix B: 14 Sentences and Summary of Results

Sentences	Number of Answers	No reply	Number of Answer Categories
1. Using my Smartphone is...	93	4	10
2. The functions of my Smartphone are...	94	3	13
3. My Smartphone is best for...	93	4	13
4. My Smartphone is not suitable for...	84	13	12
5. I think the appearance of my Smartphone is...	91	6	7
6. My Smartphone feels...	87	10	14
7. When I use my Smartphone, I feel...	91	6	12
8. I'm happy with my Smartphone, because...	93	4	14
9. The problem with my Smartphone is...	89	8	15
10. It is irritating that my Smartphone...	87	10	15
11. If other people have paid attention to my Smartphone, they...	88	9	13
12. The owner of Smartphone is typically...	78	19	11
13. In my own culture, my Smartphone...	75	22	14
14. Compared to other Smartphones, my Smartphone is...	91	6	8
Total:	1234	124	171