NATURAL LANGUAGE IN MEASURING USER EMOTIONS
A QUALITATIVE APPROACH TO QUANTITATIVE SURVEY-BASED EMOTION MEASUREMENT

Leandro Miletto Tonetto and Pieter M. A. Desmet
Department of Design, Universidade do Vale do Rio dos Sinos (Unisinos)
Zooma - Consumer Experience, Porto Alegre, Brazil
Department of Industrial Design, Delft University of Technology, Delft, the Netherlands
ltonetto@yahoo.com , P.M.A.Desmet@tudelft.nl

ABSTRACT

This paper presents an approach to developing surveys that measure user experiences with the use of natural everyday language. The common approach to develop questionnaires that measure experience is to translate theoretical factors into verbal survey items. This theory-based approach can impair the questionnaire's ecological validity because items can suffer from a disparity between the abstractness of the theory and the concreteness of actual user experience. Moreover, items formulated by researchers may not be fully understood by respondents because they do not represent their natural language. In the proposed approach the items are (a) generated by respondents instead of by researchers, and (b) are domain-specific instead of general. The approach is tested in an application study in which a survey was developed to assess the affective responses of car owners to their cars. This resulted in survey with 71 items, each representing participants' natural speech about car experiences. In an online platform, 538 people answered the survey. Reliability of data, tested with Cronbach alpha index, was 0.94, suggesting that the procedure was able to result in a survey which results are highly reliable to measure drivers' appraisals to their cars.

Keywords: Measurement of emotion; Survey Development; car experience

INTRODUCTION

Emotion has historically been an object of analysis in the social sciences. In the 90's, it also started to appear in publications of design researchers, initiating the 'Design and Emotion' research domain (for an overview, see Desmet & Hekkert, 2009). Jordan (1999) published some landmark research, developing a framework that distinguishes various sources of pleasure in human-product interaction, followed by other influential work of Desmet (2002), Norman (2004), and Hassenzahl (2010), investigating the emotional aspects of human-product and human-computer interaction.

Research in this domain has resulted in tools and techniques that support designers in designing products with the intention to evoke or avoid particular (predefined) experiences (Demir et al., 2009). Along with the growing interest in affective user experiences, the interest in instruments to measure or assess experience has also increased. These instruments can be used to understand affective responses experienced in the interaction between humans and products, and the role of the product's design in these experiences. Asking open questions is the most straightforward mean for assessing user experience. Open questions can produce detailed and subtle insights, mostly because they enable participants to freely use their own words and articulations. For this reason, they are regularly used in research-driven design approaches, such as in co-creation (Koskinen et al., 2003) and context-mapping (Sleeswijk Visser et al., 2005). A disadvantage is that open answers do not support quantitative analysis. Moreover,

administering, analyzing, and interpreting the results is typically time-consuming. For this reason, design researchers have also used instruments with a more quantitative nature, initially by borrowing them from the social sciences (for reviews see Laurans et al., 2009; Poels & De Witte, 2006). Examples are the Geneva Emotion Wheel (Scherer, 2005), a verbal self-report instrument that measures 20 distinct emotions, and the PANAS scale, which measures two independent dimensions: one for positive affect (PA) and the other for negative affect (NA), each representing a basic component of feelings (Tellegen, 1985). A third example is the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), a non-verbal (pictorial) questionnaire that measures three basic dimensions of emotions: valence (positive versus negative), arousal (calm versus excited) and dominance (or feeling of control).

Besides these (and similar) general instruments, domain-specific surveys have been developed that are sensitive to the particular characteristics of user experiences. Most experiences in the user-product relationship are felt at low intensities: products tend to elicit frustration more readily than fury, and joy more readily than ecstasy. Additionally, rather than eliciting one single experience, products may elicit multiple (mixed) experiences simultaneously, with different product aspects (e.g., general appearance, particular details, implicit and explicit expectations, and associated, remembered and imagined meanings) having different experiential impacts. Examples of instruments that have been developed to specifically measure these subtle and mixed experiences are the Product Emotion Measurement instrument, a non-verbal scale to measure emotions in response to consumer products (PrEmo; Desmet, 2003), and a scale for measuring affective responses for interactive products (Hassenzahl et al., 2010).

Within the variety of approaches, the traditional item-based questionnaire is used most often because they are easy to develop and administer (not requiring sophisticated instruments or software), and because they support quantitative investigations. Moreover, questionnaires are flexible because items can be selected to represent various types of user experiences, such as aesthetic experience (i.e., beauty), experience of meaning (e.g., elegant, fresh), or emotional experiences (e.g., contempt, desire).

The main disadvantage of item-based questionnaires is that rich experiences are reduced to a set of theory-driven abstract affect factors. These factors may not be the relevant types of experiences users typically have in response to the given product category. This may be useful for testing concepts and hypotheses, but not suitable for the early stages in the design process in which the inspirational quality of the resulting data is vital (Desmet & Schifferstein, 2012). In this paper, we therefore propose an alternative approach that aims to combine advantages of open questions and item-based questions: an item-based questionnaire that uses natural and domain-specific language. This should generate results that are both detailed and subtle, and are easy to analyze and allow for quantitative investigations. We will first discuss the general approach in questionnaire items, and then explain our approach, which involves interviewing product users. We then apply this approach to the development of a car experience survey, and test the developed survey in an online study. Possibilities and challenges of the approach are discussed in the discussion section.

DEVELOPING ITEM-BASED QUESTIONNAIRES

Although there is no clear consensus in the scientific community on how to design surveys, item generation is usually theory-based, requiring the researcher to translate theoretic factors or variables into survey items (Hinken, 1998). There is a series of guidelines on how to write survey items: statements must be simple, short and familiar to the group of respondents. Items that assess behaviors should not be mixed with those which assess affective responses. Each item should assess a single issue, avoiding combinations of issues (e.g., “comfortable and complete”), since the results will be ambiguous. Items should not represent more than one theoretical construct, to prevent respondents’ misunderstandings and invalid results. Leading questions should be avoided, and items of which it is overly obvious that all respondents will most probably answer similarly should not be used, since they tend to generate little variance. Negatively worded items should be avoided, but if used they

must be carefully worded to make sure that the respondents will understand them properly.

Although it is a common approach, translating theoretical factors into verbal survey items can impair the questionnaire's ecological validity: even though researchers can find good levels of data reliability in surveys with self-formulated items, it is questionable to what extent these surveys reflect the actual reality. Do they really uncover how people respond, think and express about particular subjects? Theory-based items can suffer from a disparity between the abstractness of the theory and the concreteness of the actual experience. For affect measurement in the design domain this problem is pungent because of the highly heterogeneous nature of the stimuli: driving a car, using spread sheet software, wearing a coat, using a mobile phone – these are all usage situations that involve very diverse kinds of experiences. In addition to this risk, theory-based items can also threaten face validity (Nevo, 1985). Items formulated by researchers may not be fully understood by respondents because they are not generated from their natural language. Are researchers sufficiently conscious of how people with different characteristics respond to any kind of stimuli? Do they know every kind of possibility on how people would respond to these stimuli?

We challenge the conventional approach to constructing item-based questionnaires by proposing an approach in which the items are (a) generated by respondents instead of by researchers, and (b) are domain-specific instead of general. In this approach, items are not theory-based and formulated in a neutral way, but interview-based and reflecting everyday language. We believe that research based on verbal statements to express consumers' experiences (e.g. using surveys) has an intrinsic qualitative character. Even when the aim is to understand large scale patterns of human-product interaction, which involves measurement, qualitative research has an important role to play. The qualitative stage enables the researcher to understand how people express about a subject and, by consequence, how they should word sentences in their surveys to create more identification from people regarding what is being measured.

NATURALIST APPROACH TO DEVELOPING ITEM-BASED USER EXPERIENCE

The naturalistic approach consists of four steps. With the first three steps, the questionnaire is developed; the fourth step is to administer the questionnaire:

- Step 1: Generating the item pool
- Step 2: Testing content validity of the items
- Step 3: Testing face validity of the questionnaire
- Step 4: Administrating survey & factor analysis

We first describe these four steps, and then illustrate them with an example in which they were used to develop a questionnaire that measures emotional experiences of car owners. Note that because the approach focuses on natural language, the questionnaire should be developed for a particular product category.

STEP 1: GENERATING ITEM POOL

The first step is to do in-depth interviews to uncover and register how people describe their feelings about the product category of interest. For this step, an extensive interview guide was developed that enables the interviewer to covers all relevant domains of emotional experience. In the development of the interview guide, the appraisal theory of user experience was used that was proposed by Demir et al. (2009). They indicated that seven general 'appraisal components' cover the variety of emotional user experiences (see Table 1). Each component represents underlying questions.

Appraisal component	**Example of related appraisal question**
Motive consistency	Does the product help me in achieving my goals?
Intrinsic pleasantness	Is the product pleasurable?
Expectation confirmation	Does the product match with my expectations?
Agency	Who (or what) is responsible for this product?
Standards conformance	Does this product match with my social norms and standards?
Certainty	How certain am i about how this product?
Coping potential	Can I handle or deal with the product?

Table 1. Seven appraisal components representing emotional user experience

In order to represent a wide overview of possible emotional experiences, the experience questionnaire should cover all seven appraisal components (Demir et al., 2009). For this reason, the interview guide covers these seven appraisal components, including a list of discussion topics for each component. A fragment from the guide can be found in Table 2. The full interview guide can be obtained on request from the authors.

Standards Conformance Component: How does this product relate to social norms and standards?
1. Is your product better, or worse than the other models on the market? 1.1 Performance (related to usage & maintenance) 1.2 Appearance 1.3 Building materials/build quality 1.4 Other aspects considered relevant by the user 2. What are the reasons for each one of the given answers?
Intrinsic Pleasantness Component: Sensorial Pleasantness
1. How does it feel to have/use your product? 2. Are there aspects that you find pleasurable? Which aspects? (Explore senses) 3. Are there aspects that you find displeasurable? Which aspects? (Explore senses)
Motive Consistency Component: Motives and Needs
1. How does this product relate to what you need, or want to do with it? 1.1 Characteristics of the product itself 1.2 Activities supported or enabled by using the product 1.3 The effect of the product on (social) identity 1.4 The role played by the product on self-expression

Table 2. Fragment from the interview guide.

Respondents involved in this step should own a product from the given category. We advise to invite people who have strong feelings to their product (either love or hate it). When recruiting respondents, they can be asked how much they agree with the sentences “I love my ….” and “I hate my ….”, using a 5-point Likert scale (ranging between completely agree to completely disagree). Select at least five respondents who completely agree to love their product, and five who completely agree to hate their product to start, or develop more interviews until the data saturation. The interviews should be appointed in moments and places of convenience for the respondent, and the product should be present. Our experience is that the interviews take 45 to 75 minutes per respondent. Interviews can be registered with video and transcribed. With content analysis, all affect-related sentences can be identified and subsequently categorized on the basis of similarity.

Then next step is to select the most clear and direct sentence from each category to serve as the basis for the survey’s items.

STEP 2: TESTING CONTENT VALIDITY

The second step is to invite an expert to evaluate the generated items in terms of relevance and formulation. The expert should be experienced in doing research related to the particular product category. This step can be considered an expert-based content validation, in which the expert determines the extent to which the item pool represents all facets of a given usage experience. The expert can suggest minor modifications of formulations, advise to select different sentences per category, or to split categories in relevant sub-categories. Note however, that the original sentences should be adapted as little as possible, to preserve the original user’s natural expression.

STEP 3: TESTING FACE VALIDITY

In step 3, the actual questionnaire is constructed by representing each item with a Likert-type 5-point scale that measures the participant’s level of agreement with the sentence. The scale is anchored in its extreme values (i.e. completely disagree; completely agree). The order of sentences should be randomized and additional questions (e.g. about demographics) can be added. The face validity of the survey should be tested in a pilot study. The validity of a test at face value assesses whether the test "looks valid" to the untrained observers. In this case, it is to be used to test the adequacy of the language and item presentation. Respondents fill out the questionnaire, and then are interviewed to assess if they understood all items and felt comfortable with all formulations.

STEP 4: ADMINISTRATING SURVEY & FACTOR ANALYSIS

The fourth step is to administer the survey.

APPLICATION FOR ASSESSING EXPERIENCES OF CAR OWNERS

DEVELOPING THE QUESTIONNAIRE

For the first development step, ten respondents (5 females; 18 to 54 years old; 3 undergraduates, 4 professionals, and 3 graduates) were recruited

through social networks. Five respondents loved their cars, and five hated their cars. Using the guide described above, respondents were interview individually inside their cars.

The interviews were analysed through the content analysis technique, from which all the answers provided (over 120 pages) were classified in 71 relevant categories of answers. Only the answers that were actually related to the research aims were selected and grouped by content/meaning.

Each one of the 71 categories of answers related to at least one of the seven appraisal components. The most clear and direct sentences to express each one of the 71 categories were selected, resulting in 71 sentences, which were selected to inspire the survey's items (see the fourth column in Table 4). These sentences were adapted as little as possible, to keep the original driver's vocabulary. Some changes were made to make the sentences make sense out of context of the conversation (i.e., the full transcription of the interview) and to eliminate slang.

For the second step, a professor of design aesthetics with experience in measuring user experience in the domain of car experience participated as the expert. In a meeting with the authors, the expert discussed the relevance and form of all and each one of the items and suggested several modifications, mostly related to dropping emotional words (focusing on appraisals to love or hate).

We used an online platform to build the survey. The procedure started with a set of questions regarding the respondent's demographics and his or her car's characterization. Next, the respondent answered his or her level of agreement with each of the 71 sentences. The data was automatically registered in the online platform, to be downloaded when needed. The generated databank was created in a Microsoft Excel format to be converted to the SPSS platform.

In step 3, two groups of 34 and 27 undergraduate students from a southern Brazilian University filled out the questionnaire. They first answered the survey in a media lab, and then discussed each item of it, suggesting insignificant changes. The only claim was that the survey took "a bit long" to be answered: between five and ten minutes. Given these results, no further modifications to the questionnaire were made.

APPLYING THE QUESTIONNAIRE

Respondents

In total, 710 respondents answered the survey. 75.8% (N=538) completed the whole survey without leaving questions unanswered. The final analysis, presented in this section, was done considering these 538 participants. Mean age was 31.19 (Std=10.94), 56.7% were females and 43.3% were males. Regarding their academic background, 30.1% were undergraduate students, 22.7% were professionals holding a bachelor degree or equivalent and 38.5% were graduated. Their car brands were mainly Chevrolet (16.2%), Volkswagen (15.8%), Fiat (14.9%) and Ford (14.3%). These cars were bought either new (65.2%) or used (34.8%). About how they use their cars, 76.9% of the sample does a daily use of it, and 51.7% is a single user, while 24.5% is the main user, but not the only one.

Results

Reliability of data was tested with Cronbach alpha index. The result for the 71-item scale was 0.94. The KMO (Kaiser-Meyer-Olkin) was used in order to measure the sampling adequacy for a satisfactory factor analysis. The result was 0.934, which suggests that our data was suitable for factor analysis (Kaiser, 1974). In addition, we used the Bartlett's test of sphericity. It tests whether the correlation matrix is an identity matrix, which would indicate that the factor model is inappropriate. Its result was 16664.814. This indicates that the factorial model is adequate to be used (Norusis, 1994). The communalities presented a mean of 0.589, indicating that each variable shares a good amount of variance in the analysis with the other measured variables. All items in the scale have communality equal or above 0.462.

The factorial extraction suggested 15 factors, and the lowest result regarding a factor explained 1.72 of the total variance. When the results suggest more than ten factors, it is usual that the eigenvalues are below 1.0, which would indicate that we should not accept this number of factors. This was not the case

regarding this survey, since all eigenvalues are above 1.0. It was found that the 15 factors explain 58.989% of the total variance from the group of variables that are part of the instrument. Even though the variance explained by each factor is relatively balanced among them, the distance between the first (23.318) and the second (7.185) factors was higher. The Varimax orthogonal rotation showed a better distribution of the variance among factors. These results can be seen in Table 3. The items' distribution in factors, considering their factorial extraction observed after the Varimax rotation and the Kaiser normalization, can be seen on Table 4.

Factors	Initial Extraction			After rotation		
	Eigenvalues	% Variance	% Cumulative Variance	Eigenvalues	% Variance	% Cumulative Variance
1	16.555	23.318	23.318	8.356	11.769	11.769
2	5.101	7.185	30.502	7.957	11.207	22.976
3	3.650	5.141	35.643	4.975	7.007	29.983
4	2.091	2.945	38.588	2.227	3.137	33.120
5	1.948	2.744	41.332	2.167	3.052	36.173
6	1.674	2.358	43.690	2.114	2.977	39.150
7	1.432	2.017	45.707	2.048	2.884	42.034
8	1.399	1.971	47.678	1.959	2.759	44.793
9	1.300	1.832	49.509	1.707	2.405	47.198
10	1.280	1.803	51.312	1.595	2.247	49.445
11	1.187	1.672	52.985	1.465	2.063	51.508
12	1.112	1.566	54.551	1.416	1.994	53.502
13	1.079	1.519	56.070	1.350	1.901	55.403
14	1.048	1.476	57.546	1.325	1.867	57.269
15	1.025	1.443	58.989	1.221	1.720	58.989

Table 3. Factor analysis results before and after the rotation

Fac-tor	Fact. charge	Item	Items	Factor description
1	0.790	66	It already gave me so many problems, that I'm afraid of it.	
	0.746	43	This car just gave me problems.	
	0.739	65	I go through situations with my car, caused by it, that make me angry.	
	0.717	71	My car is a Pandora's box. Anything can happen to it.	
	0.705	47	I am traumatised with this car.	
	0.703	70	I need to be conscious to drive my car, due to the problems I have with it.	
	0.660	67	I don't feel safe with my car, because it fails and I don't know the reason.	
	0.585	35	I thought it was going to give less problems than it gave.	
	0.579	63	The impression I have is that something will break in the car.	
	0.550	33	It never lets me down.	Lack of reliability and safety
	0.526	40	I have a confidence in it, because it never caused any problems.	
	0.521	29	It makes unbearable noises.	
	0.517	68	My car makes me embarrassed sometimes.	
	0.483	59	I believe it's not going to give many problems, I think it's going to be very durable.	
	0.479	56	It doesn't give me any surprises, like things that fail or fail slowly.	
	0.426	62	I have confidence in the car, that if something goes wrong, I can get good advice.	
	0.423	38	It obeys the orders as I want it to.	
	0.269	18	I don't feel well when I look at it.	

Table 4. Factors, its items and the rotated component matrix

Fac-tor	Fact. charge	Item	Items	Factor description
2	0.766	2	It seems like it's a part of me.	Personal identity
	0.735	4	I identify myself with my car, since I have had many happy times with it.	
	0.734	1	My car suits me.	
	0.704	6	My car says something about me.	
	0.688	10	It gives me the feeling of freedom, more than other cars.	
	0.687	7	It's able to blend itself to my way of life.	
	0.657	26	The other cars are not as fun as mine.	
	0.573	8	I'm a fan of my car; it allows me to do everything I want to do with it.	
	0.552	42	I feel more confident to drive it, than I felt before with other cars.	
	0.546	17	I like everything about it.	
	0.536	46	There's no better car than this.	
	0.527	36	It seems that I got to drive better, after I got this car.	
	0.511	5	It's perfect for me, there are no problems.	
	0.511	9	It's a dream car.	
	0.449	37	It's a car that I have adapted myself very well to.	
	0.412	34	It represents everything that I expected from it.	
	0.403	49	It was more valuable to buy this car, than any other car in its category.	
	0.377	16	There are several advantages that convinced me to buy the car.	
	0.272	25	I see other cars and I don't feel jealous of them.	
3	0.657	21	Its engine is quiet.	Comfort due to the lack of sensorial displeasure
	0.634	22	It doesn't have internal noise.	
	0.606	20	My car is comfortable.	
	0.538	50	It has everything I need, it's really complete.	
	0.533	28	Its internal parts are good to touch.	
	0.496	64	Even in adverse situations, like traffic jams, I like to stay in the car, due to the comfort it offers.	
	0.493	23	I like its internal space.	
	0.488	19	It smells/smelled new, really good.	
4	0.758	41	There is something in my car's structure that seems fragile.	Fragility related to the car´s power
	0.741	55	I feel that it's fragile, because it has many plastic components.	
	0.363	14	My car is more resistant than others.	
5	0.566	39	I don't have any technical difficulties, everything is self-explanatory.	Ease of use
	0.382	52	Its performance is better than other cars in the same category.	
	0.374	27	It's really easy to adjust my car to my needs.	
	0.311	3	I get to do all I want to with my car.	
	0.251	11	It's easy to park.	
6	0.693	30	There is a respect from other people through this car being bigger.	Social power
	0.556	31	My car is robust.	
7	0.729	61	Its maintenance is expensive.	Financial factor (maintenance)
	0.679	58	I don't believe the maintenance will be too expensive.	
	0.530	60	There is a lack of parts, when I need to find them.	
8	0.757	12	One of the reasons I bought this car because it spends almost nothing in fuel.	Financial factor (fuel)
	0.753	51	It spends very little in fuel, compared to other cars in its category.	
9	0.765	45	Its insurance isn't so expensive.	Financial factor (taxes and insurance)
	0.709	44	The annual road tax isn't so expensive.	
	0.521	13	It's not targeted in theft.	
10	0.665	54	It doesn't have the same things that other cars have in its category.	Inferiority
	0.454	32	Its design is a dated.	
	0.398	53	I don't think this car is as comfortable as others I have seen.	
11	0.507	57	I know what's going to happen to the car in the future; I won't have any surprises.	Certainty
12	0.586	15	I believe it should be easy to resell my car, if I wanted to.	Reselling potential
13	0.654	48	For the paid price, it should have been better made.	Expectation disconfirmation
14	0.800	69	The biggest problem of this car is the driver.	The driver is the problem (not the car)
15	0.794	24	It does not have areas that you cannot see.	Visual wideness

Table 4. Factors, its items and the rotated component matrix (Continuation)

Even though the factors from 11 to 15 are composed by only one item each, we decided to keep them in the survey, due to the importance of their contents and because they do not seem to share a lot in meaning with any other general factor. Some of these last 5 items/factors even have some of the highest communality scores in the 71-item survey. Item 24 (factor 15) has the 4th highest score (0.688), 69 (factor 14) has the 11th highest communality (0.668), while item 48 (factor 13) in is the 20th position (0.632). Even the other two (factor 11, item 57, and factor 12, item 15) do not have low communality scores: 0.487 and 0.680, respectively. Each one of the 15 factors was named according to its items contents; see the last column in Table 4.

Convergent validity, the degree to which a factor converges on the global scale results, was tested with a multiple linear regression, with the items global mean score as dependent variable, and the mean score for all items in each factor as independent variables. This analysis found a significant Adjusted R Square (0.998), indicating that most of the variance of the global results is explained by the 15 factors that are part of the scale. Table 5 shows how much each factor predicts the scale global results.

Factors	Unstandardized Coefficients	Standardized Coefficients	T
Const.	0.010		1.316[a]
1	0.261	0.338	132.962[b]
2	0.273	0.404	145.279[b]
3	0.099	0.168	57.771[b]
4	0.040	0.069	32.533[b]
5	0.071	0.095	37.779[b]
6	0.028	0.060	27.657[b]
7	0.046	0.081	38.323[b]
8	0.026	0.056	28.209[b]
9	0.041	0.071	37.057[b]
10	0.044	0.075	35.986[b]
11	0.015	0.035	18.586[b]
12	0.014	0.029	14.727[b]
13	0.012	0.028	14.141[b]
14	0.014	0.033	18.201[b]
15	0.014	0.034	18.377[b]

[a] Significant differences at α=0.05; [b] α=0.001

Table 5. Regression analysis

In Table 5, it can be observed that the highest results are 0,404 (factor 2) and 0,338 (factor 1), both of them considerably distant from all the others.

DISCUSSION AND CONCLUSION

In this paper we introduced an approach to develop quantitative questionnaires with the use of language that people naturally use to express their feelings about a particular product category. The main feature of the approach is that the questionnaire items are created on the basis of interviews instead of on the basis of theoretical constructs. The approach was applied in the development of a survey that measures car experiences with 71 natural statements. An application study with 538 respondents indicated that these 71 statements represent 15 underlying experience-factors. This amount is interesting because regular (theory-based) questionnaires generally result in two up to six meaningful factors. This indicates that the use of natural language enables the researchers to obtain more nuanced and rich insights in user experiences than conventional quantitative experience measurements.

A possible explanation is the holistic nature of user experience. Most regular experience questionnaires measure a particular type of affect, such as emotions. Strictly speaking, the concept of emotion refers to a particular and specific affective phenomenon: a relatively brief episode of coordinated brain, autonomic, and behavioral changes that facilitate a response to an external or internal event of significance for the organism (Scherer & Peper, 2001). In the design (research) discipline however, the word emotion is used to represent a wider perspective, including all kinds of affective aspects involved in the user–product relationship (see Desmet et al., 2008). As a consequence, the items in regular questionnaires usually only represent part of the actual user experiences. With items that are actual expressions of experience, the naturalistic survey represents people's natural way of communicating how they feel about the particular product category.

It is also important to highlight that the high number of factors identified in this approach might also be due to the fact that the study was domain-specific, focusing on affective responses specifically from car owners.

In our view, we should never put the need of measurement in front of the need of understanding human phenomena. The approach described in this paper demonstrates that regular rules for creating items (e.g. simplicity, clarity, etc.) are not poured in concrete and should not be considered a sufficient checklist for creating valid questionnaires. When integrating a naturalistic – qualitative – dimension to a survey, researchers need to be flexible in applying general rules for formulating items. For example, people sometimes combine rather than separate experience attributes in their expression of feelings. It then seems unnatural to split them up in separate items because that is advocated by the rule. If we do not fixate on "item purity", we can gain ecological validity. For example, the car survey included the expression "I have a confidence in it, because it never caused any problems." As an item, this sentence is flawed because it includes a causal relationship. Hence, one can decide to split the sentence into two items: "I have confidence in it" and "It never causes problems." In the interviews however, participant's always justified their (lack of) confidence in their cars. A statement about confidence does not make sense if isolated from the reason. Although the combined sentence may not be customary in measurement techniques, it does represent everyday car experiences. Note that the development of dedicated surveys in which all items are sentences that are directly related to the consumer's experience represents a trend in psychometry, which is to develop specific instead of general surveys.

The naturalistic approach has advantages and disadvantages. It gains in specificity and nuance in the data, but it also requires more time and effort to develop the questionnaire. For each product category, a new questionnaire needs to be developed. Moreover, cultural differences may pose a challenge. There are differences to be expected between cultures in how people express their feelings towards a product category. This will complicate cross-cultural studies because these will require the development of different surveys for the different cultures, and the results will be difficult to compare because of the non-standardized measurement.

Even though the aim of this paper was not to discuss the links between results from factor analisis and appraisal components, a brief description on each factor can be found below, followed by their relations with the relevant appraisal components (between brackets).

- Factor 1: Lack of reliability and safety (certainty and coping potential);
- Factor 2: Personal identity (motive consistency);
- Factor 3: Comfort due to the lack of sensorial displeasure (intrinsic pleasantness);
- Factor 4: Fragility (motive consistency) related to the car´s power; opposed to the social power included in the 6th factor;
- Factor 5: Ease of use (mostly motive consistency);
- Factor 6: Social power (related to motive consistency);
- Factor 7: Financial factor – maintenance (motive consistency);
- Factor 8 Financial factor – fuel (motive consistency);
- Factor 9: Financial factor – taxes and insurance(motive consistency);
- Factor 10: Inferiority (related to standards conformance);
- Factor 11: Certainty (certainty);
- Factor 12: Reselling potential (related to motive consistency, but as a complimentary car characteristic, not as a motive by itself);
- Factor 13: Expectation disconfirmation (expectation confirmation);
- Factor 14: The driver is the problem, since the car has no problems (agency and coping potential);
- Factor 15: Visual wideness (motive consistency, even tough its content can suggest that it is related to sensorial pleasantness).

Motive consistency seemed to be related to almost all factors mentioned above. Appraisals can be understood sometimes as part of other appraisals, since one can buy a new car in order to feel more comfortable when driving, e.g. It is related to motive

consistency, but it is also about sensorial pleasantness.

As a next step in further developing and refining the naturalistic approach, we plan to conduct a comparison study in which we compare outcomes of the naturalistic questionnaire with those of conventional experience questionnaires. Additionally, the interview guide (used in the first step of the approach) will be refined by testing it in several product domains with a variety of experience researchers.

ACKNOWLEDGEMENTS

The first author would like to thank the Department of Industrial Design at Delft University of Technology for hosting him as post-doctoral visiting research fellow (2010-2011) to work on this research.

This research was supported by the MAGW VIDI grant number 452-10-011 of The Netherlands Organization for Scientific Research (N.W.O.) awarded to P.M.A. Desmet.

REFERENCES

Bradley, M.M., Lang, P.J. (1994) Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, Vol.25, Issue 1, 49-59

Demir, E., Desmet, P.M.A., Hekkert, P. (2009) Appraisal Patterns of Emotions in Human-Product Interaction. *International Journal of Design*, Vol. 3, Issue 2, 41-51

Desmet, P.M.A. (2002). *Designing emotions*. Unpublished doctoral dissertation, Delft University of Technology, Delft, The Netherlands.

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, P.C. Wright (Eds.). *Funology: from Usability to Enjoyment*. Dordrecht: Kluwer Academic Publishers, 111-123

Desmet, P.M.A., Hekkert, P. (2009) A decade of design and emotion. *International Journal of Design*, Vol. 3, Issue 2, 1-6

Desmet, P.M.A. Schifferstein, N.J.H. (2012). Emotion research as input for product design. In J. Beckley, D., Paredes, & K. Lopetcharat (Eds.), *Product Innovation Toolbox: A Field Guide to Consumer Understanding and Research*. Hoboken, NJ: John Wiley & Sons, Inc, 149-175

Desmet, P.M.A., Van Erp, J., Karlsson, M.A. (2008) Design & Emotion Moves; Introduction. In: P.M.A. Desmet, J. Van Erp, & M.A. Karlsson (Eds.). *Design & Emotion Moves.* Cambridge: Cambridge Scholars Publishing, 1-15

Hassenzahl, M. (2010) *Experience Design: Technology for all the right reasons*. San Francisco: Morgan Claypool.

Hassenzahl, M., Diefenbach, S., Göritz, A.S. (2010) Needs, affect, interactive products – Facets of user experience. *Interacting with Computers*, Vol. 22, 353-362

Hinken, T. (1998) A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, Vol. 1, Issue 1, 104-121

Jordan, P.W. (1999) Pleasure with products: Human factors for body, mind and soul. In W.S. Green & P.W. Jordan (Eds.), *Human factors in product design: Current practice and future trends*. London: Taylor & Francis, 206-217

Kaiser, H. (1974) An index of factorial simplicity. *Psychometrika*, Vol. 39, 31-36.

Koskinen, I., Battarbee, K., Mattelmäki, T. (2003). *Empathic Design, User experience in Product Design*. Helsinki: IT Press.

Laurans, G., Desmet, P.M.A., Hekkert, P. (2009) Assessing Emotion in Interaction: Some Problems and a New Approach. In: A. Guenand (Ed.), *Proceedings of the 4th International conference on Designing Pleasurable Products and Interfaces*. Compiegne (France), 13-16 October 2009.

Nevo, B. (1985) Face Validity Revisited. *Journal of Educational Measurement,* Vol. 22, 287-293

Norman, D. (2004) *Emotional design: Why we love (or hate) everyday things*. New York: Basic Books.

Norusis, M. (1994) *SPSS professional statistic 6,1*. Chicago: SPSS.

Poels, K., Dewitte, S. (2006) How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *Journal of Advertising Research*, Vol. 46, Issue 1, 18-37

Scherer, K.R. (2005) What are emotions? And how can they be measured? *Social Science Information*, Vol. 44, Issue 4, 693-727

Scherer, K.R., Peper, M. (2001) Psychological theories of emotion and neuropsychological research. In F. Boller & J. Grafman (Eds.). *Handbook of Neuropsychology*. Vol. 5 (Emotional behaviour and its disorders, Ed. G. Gainotti). Amsterdam: Elsevier, 17-48

Sleeswijk Visser, F., Stappers, P.J., van der Lugt, R., Sanders, E.B.N. (2005) *Contextmapping: Experiences from practice*. CoDesign, Vol. 1, Issue 2, 119-149

Tellegen, A. (1985) Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A.H. Tuma & J.D. Maser (Eds.). *Anxiety and the anxiety disorders*. Hillsdale, NJ: Erlbaum, 681-706