The Catalog Usability Questionnaire – Adoption and Validation of a Usability Scale for Print-Catalogs

Sebastian Schulz, Gunnar Mau, and Günter Silberer
Institute of Marketing and Retailing, University of Göttingen, Germany
sschulz8@uni-goettingen.de
gmau@uni-goettingen.de
gsilber@uni-goettingen.de

Abstract: The usability of interfaces is mainly tested via standardized-questionnaires or via less standardized process-oriented verbal protocols. Standardized questionnaires have got the advantage of generating reasonable indices with quite low effort. In contrast, process oriented verbal protocols offer the opportunity of gaining more thorough insights into the interaction and the problems that might come up.

While efficient standardized methods for the measurement of online media with their non-linear structure already exist, there are no such methods to measure the usability of linear structured offline media. However, it can be assumed that usability of offline print media plays an important role for a brand's image and the buying behavior of customers.

The objective of our study was to adapt a scale to measure usability of online shops to print catalogs in order to measure their usability. A field study (n = 60) and a laboratory experiment (n = 92) have been made to test the reliability and validity of the adapted scale. Participants were asked to evaluate the catalogs of two travel companies by using the adapted questionnaire.

Concerning the reliability, in both studies the scale showed good internal consistency. The values of Cronbach's alpha were comparable to results in studies using the original scale with online shops.

Concerning the validity, results show correlations between usability and buyer's intention as well as recommendation. Furthermore, the usability scores of both catalogs differed as experts (n = 20) expected. Moreover, the questionnaire was validated using the results of verbal protocols that were collected simultaneously in the laboratory study. In example, the correlation between usability scores and problems that came up during the interaction was proven.

Overall, the results support the usefulness of the adapted scale to measure the usability of print-catalogs. Implications for research and management will be discussed.

Keywords: usability scale implementation, verbal protocols, print-catalogs

1. Introduction

If potential customers decide to purchase a trip or holidays, they turn mostly to print catalogs as a source of information (cf. Fodness & Murray 1999). If the decision to buy is based to a large extent upon information from the catalog, its design and, above all, usability appear to be significant. After all, only the information the customer can find in the catalog can be considered in the purchase decision.

In the study presented here, we concentrated on travel catalogs as the catalog plays a particularly significant role in this industry. The catalog serves as a permanent reference book for customers. Furthermore, it is significant for online commerce as nearly 40 percent of online buyers first use the dealer's catalog prior to ordering on the Internet (cf. Bhv 2005).

Anyone who is searching for instruments to measure the usability of print catalogs in scientific publications will be disappointed. Hodgson (1993: 51) pointed out that the scientific research on usability in the print field primarily concentrates on newspapers, advertisements, magazines and pocket books (similarly Schmeißer et al. 2005). Only a few publications have addressed the topic of catalogs or travel catalogs at all and only a small proportion of these the effect of texts and other elements in catalogs and tourist brochures (e.g. Hodgson 1991, 1993; Goossens 1994). The referral to in-depth interviews and focus groups as evaluation instruments, as are often used, appears to be as time-consuming as it is costly. Evaluation methods which work with verbal protocols seem far more promising. The low level of standardization may cause various problems (e.g. in the coding and evaluation of data) but it offers the opportunity of gaining more thorough insights into the interaction. A standardized interview instrument is not yet available.

In contrast, there are many studies on the topic of the usability of web sites and online catalogs and, furthermore, also on tourist web sites. Meanwhile, science and management can fall back upon validated

Reference this paper as:
Schulz S., Mau G., and Silberer G. (2007) "The Catalog Usability Questionnaire – Adoption and Validation of a Usability Scale for Print-Catalogs." *The Electronic Journal of Business Research Methods* Volume 5 Issue 2, pp 93 - 104, available online at www.ejbrm.com

instruments for the measurement of usability in this field. The use of standardized usability questionnaires has established itself as both a simple and inexpensive evaluation method (e.g. Konradt et al. 2003).

After years of the transfer of instruments and methods of offline market research in an online context, a validated online usability questionnaire (UFOS Scale, Konradt et al. 2003) is tranferred to the evaulation of the usability of two travel catalogs in the scope of this empirical study. In doing so, the use of such a survey instrument will be presented and, furthermore, the quality of the instrument and its impact on final factors will form the focus of this article. In addition, there will be a comparison between verbal protocols and the scale and implications for a combined use.

2. Usability and studies on catalogs and brochures

2.1 Results of the usability research

In recent years, the topic of usability has been addressed in many scientific and management publications. Usability primarily shifted into the focus of science and practice through the spread of modern information technology and questions on the operability of computer applications and web sites. However, usability refers to more than merely software and web site applications, as Nielsen's (1993) definition suggests: "usability is the measure of the quality of the user experience when interacting with something - whether a Web site, a traditional software application, or any other device the user can operate in some way or another". In addition, usability does not represent a one-dimensional construct. In science and practice, the usability understanding of the international norm ISO 9241(ISO 1998) has prevailed. According to this understanding, an object or application is user-friendly if its aims can be achieved effectively, efficiently and satisfying. The three criteria effectiveness, efficiency and satisfaction should also be regarded for the usability of a travel catalog as dimensions of usability.

Table 1: Categorization of usability evaluation methods (Wilhelm & Yom 2004; Stoessel 2002; Sweeney et al. 1999; Mack & Nielsen 1994; Nielsen 1994)

Theory-based methods	Expert-based methods (analytical)	User-based methods (empirical)
Cognitive Models & Architecture GOMS (Goals, Operators, Methods and Selection Rules)	Heuristic Evaluation Cognitive Walkthrough Check lists	**Surveys** Verbal Protocols Concurrent verbalization **Retrospective verbal reports** Eye Tracking

In the field of usability evaluation methods, different categorizations have developed. One which applies the most is the breakdown into theory-based, expert-based (analytical) and user-based (empirical) evaluation (cf. Sweeney et al. 1993; cf. Mack & Nielsen) (see table 1). For example, expert-orientated processes would be the cognitive walkthrough or the heuristic evaluation; all forms of survey represent user-based processes. However, Hodgson (1991) only distinguishes qualitative (e.g. in-depth interviews, face-to-face interviews) and quantitative (e.g. standardized survey) methods to investigate usability. In this study, two methods of the user-based processes are used: firstly, survey and, secondly, retrospective verbal protocols.

2.2 State of the research on catalogs, travel catalogs and brochures

Fodness and Murray (1999) examined whether the travel catalog plays a role at all in the search for information. In their study, they develop a "Tourist Information Search Behavior" model and test it empirically. In the process, seven different groups, each with different information search behavior, are identified. The results show that catalogs and brochures are almost referred to as an information source in a decision. Five out of seven groups refer to catalogs in the information search.

An initial study on catalog purchasing behavior and the influence of situative determinants was published in 1974 in the Journal of Marketing (cf. Reynolds 1974). In an experimental study, the main influencial factors were examined for purchasing with catalogs. For this purpose, the author investigated nine factors using a factor analysis from 107 items, which he used as a basis for comparison. The results show that regular catalog purchases are motivated by the need for simplicity and convenience. Summarizing, the author states that "…it appears that catalog buying is more a result of catalog offerings, as revealed in the negative attitudes toward local shopping conditions expressed by young, affluent risk takers who comprised the frequent buying group." (ibid.). Even if the results of this study reveal important findings regarding the

situative determinants of catalog behavior, they also demonstrate the singular role of the catalog content. In this respect, it goes without saying that a sensible presentation and simple usability are of great importance but this was not considered in Reynolds study.

Goossens (1994) goes into the effect of elements in the catalog more specifically and examines the effect of texts and pictures in travel catalogs on the external information behavior in an experimental study. In doing so, the author varies the size of the pictures within a 2x2 factorial design and the emotional character of the texts in a travel catalog. In each case, 5,000 copies were sent to a total of 20,000 households which had been chosen at random. The author does not determine any differences dependent on the stimuli. He attributes this to the fact that the variations of the catalog are not perceived as different (cf. Goossens 1994). Recording the usability and the the effect of the elements through a standardized questionnaire would also have been possible here but was not used.

Most notably, Hodgson dealt with "Tour Brochure Design" from the beginning of to the mid-90s. In the process, the author reports that only a small proportion of companies make the sufficient effort and utilize enough financial resources to design their catalogs in a user-friendly fashion (cf. Hodgson 1993). As, in his opinion, usability is a central factor in the success of catalogs, evaluation should be a permanent feature of catalog development. Therefore, Hodgson suggests first and foremost qualitative methods (focus groups, in-depth interviews), which he regards as "more sensitive" (cf. Hodgson 1991). However, as the cost and efforts are often very high in qualitative methods, he also quite rightly addresses quantitative methods, such as standardized questionnaires. These are admittedly less explorative and sensitive but with the appropriate effort can provide useful indications as to the problem areas.

The literature research carried out shows that the scientific research deals with a wide variety of problems from the catalog sector. However, qualitative methods dominate the evaluation methods. A standardized tool for the effective and efficient examination of the usability in due consideration of the cost-use relationship has not yet been developed.

3. Empirical Study

Within the scope of our study, the adoption of a standardized online usability questionnaire (UFOS scale) to print catalogs was tested. Furthermore, the combination with verbal protocols was examined. This study was carried out in collaboration with a large German travel agent. From a methodical point of view, the following objectives and research questions underlie the study:

- How can the standardized questionnaire for online evaluation be adapted for print catalogs?
- Does the scale fulfill the quality requirement of reliability, i.e. does it produce reliable values?
- Does the questionnaire facilitate valid measurements?
- Do the questionnaires and the verbal protocols produce the same result?
- What other results can be obtained from the verbal protocols in the evaluation of a print catalog?

Firstly, the UFOS scale, which represents the basis for the instrument developed here, will be presented below and the adaptation to the travel catalog sector elucidated. The presentation of verbal protocols and their application in the field of catalog evaluation ensues in section 3.2. This is then followed by the design of the study in section 3.3 and the presentation of the central results in section 3.4 and 3.5. The section closes with a summary.

3.1 A questionnaire to measure the usability of print catalogs

The UFOS scale developed by Konradt, Wandke, Balazs and Christophersen (2003) forms the basis for the questionnaire used here. The underlying definition of usability comprises the three dimensions, effectiveness, efficiency and satisfaction, addressed in section two (Konradt et al. 2003; ISO 9241). In the development of the questionnaire, the scientists conducted extensive research using existing usability questionnaires (e.g., SUMI, PUTQ; Kirakowski 1996) and considered these and the results obtained. Furthermore, they conducted an extensive validation of the instrument in an online context. Primarily due to the integration of different questionnaires, the systematic derivation and the available validation, we opted for the UFOS scale as a basis for the questionnaire to measure the usability of print catalogs.

The original questionnaire consists of 47 questions, which are structured into seven categories (general usability impression: 14 questions, information on the basic conditions: 7 questions, product search: 7 questions, cart handling: seven questions, product overview: 4 questions, self-description ability: 4

questions, product description: 4 questions). The original items were adjusted to the evaluation in the print catalog field. Items which did not include relevant aspects for catalog usability were eliminated. After the modification and adaptation of the original questionnaire, a total of 23 questions in five categories remained. According to expert opinion, these cover the relevant aspects of the usability of travel catalogs. The questions used and the corresponding categories are illustrated in table 2 (Catalog-Usability-Questionnaire, CUQ). The emphasis is on the category general usability impression, from which 13 of the 17 questions were adapted and adopted, and the category club overviews, which largely consists of items from the former category product overview. The category cart handling was dispensed with in the modified questionnaire.

Table 2: Catalog-Usability-Questionnaire (CUQ)

Scale: In each case, 5-Point Rating Scales from “strongly agree“ to “strongly disagree“
1. General Usability Impact (13 Items)
- This catalog makes the search for a club vacation / hotel easier
- I find using the catalog too complicated
- You can quickly gain an overview of the catalog
- Dealing with the catalog is easy to learn
- The catalog encourages me to explore it further
- The search for a trip / holiday can be completed swiftly
- The catalog offers useful advice on its use
- It is easy to find your way around the catalog
- The catalog is constructed consistently
- The catalog is well-structured
- I can always recognize important information easily
- I can always find the information which is important to me easily
- The design helps in the use of the catalog
2. Additional Information on the Travel Conditions (1 Item)
- I always have an overview of the costs which arise
3. Search for Clubs / Hotels (1 Item)
- It is easy to search for a suitable club / hotel for me
4. Club / Hotel Overviews (7 Items)
- The overview pages of the hotels / clubs are clear
- The categories of hotel / club overviews offered are selected sensibly
- The symbols on the overview pages are clear
- I always have an overview about the price of the club / hotel
- I can verify whether the club corresponds to my expectations
- I am satisfied with the descriptions of the club / hotel
- I find the pictures used helpful
5. Self-description (1 Item)
- I do not require any explanations to use the catalog

[a] Items original in German language

3.2 Verbal protocols to measure the usability of print catalogs

Besides the modified questionnaires, a further user-based method of evaluation was used within the scope of the evaluation in the form of verbal protocols (cf, section 2). Verbal protocols offer the possibility of a *process-associated* assessment of cognitions (Ericson & Simon 1993). The concurrent verbalization can be differentiated from the retrospective verbal protocols (ibid.). In the case of the concurrent verbalization, the participant expresses all thoughts on usability during the interaction with the catalog; for the retrospective verbal protocols, the participant is asked to comment on his thought process in hindsight (Büttner & Silberer 2007). The retrospective video-cued thought protocols represent a development on retrospective verbal protocols. In the process, the subsequent verbalization is supported by a video of the interaction process which was recorded during the actual activity (Silberer 2006). The methods can be distinguished through their proximity to the process, without binding cognitive capacities through verbalization during the interaction.

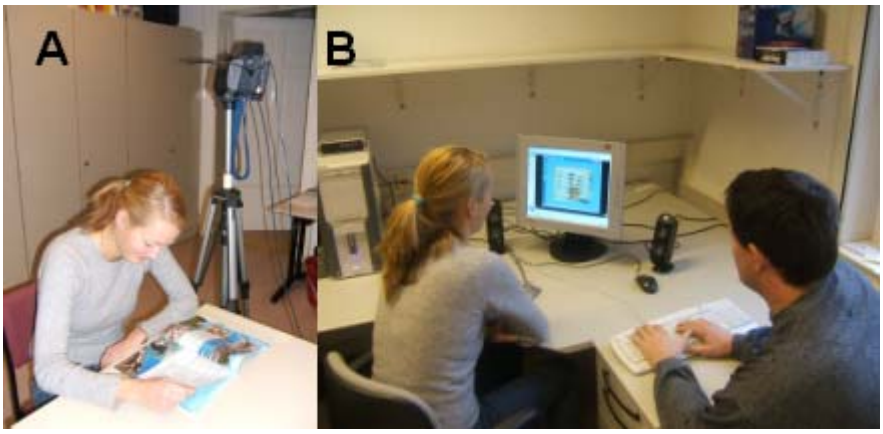

Figure 1: Test procedure in the retrospective video-cued thought protocols

In the assessment of the cognitions on usability, first the participant is filmed during the interaction with the catalog (see fig. 1 A). Afterwards, the video is shown to the participant on the PC and all thoughts which he/she expresses on usability are recorded (see fig. 1 B). A combined file is subsequently put together from the video and the commentary.

The videos produced facilitate both a qualitative and quantitative assessment of the usability problems expressed. For the quantitative assessment, the participant's comments have to be classified in a category system.

The category system used here was established from literature and expert discussions with practitioners (Nielsen 1993). For an objective classification of the positive and negative statements, it should comprise a manageable number of selective categories. The following category system was used:

Statements on:

- Category 1: impression and expressiveness of the pictures and photos
- Category 2: clarity and design of the colors
- Category 3: general color scheme
- Category 4: information content of the text
- Category 5: overall composition of the catalog

3.3 Design of the study and operationalization

The test design comprised a preliminary and a main study, which in turn consisted of three sub-studies. Within the context of the preliminary study, both catalogs were evaluated by experts with regard to different usability dimensions in order to be able to validate later results. In doing so, the experts had to gage the overall usability, effectiveness, efficiency, satisfaction and the 5 main-categories of the usability questionnaire (general usability impression, additional information, search for clubs, club overviews, self-description ability) on 5-point rating scales.

The main study comprised a laboratory study with $n1$ = 92 test participants, a field study with $n2$ = 60 participants and a final expert survey with $n3$ = 20 participants. The laboratory study served the questionnaire test under controlled conditions and the gathering of the retrospective video-cued thought protocols. With the second study, users of catalogs in three travel agencies were interviewed. The final expert survey served to confirm the results found on the basis of the Catalog-Usability-Questionnaire.

Figure 2: Catalogs used in the study. Cover and two pages (Aldiana GmbH and Robinson-Club GmbH)

The Catalog-Usability-Questionnaire was used in all three sub-studies. The procedure was also practically identical in all three sub-studies: after the measurement of different control measures, such as experiences and socio-demography, in a pre-questionnaire, the participants familiarized themselves with a travel catalog for about 12 minutes (see fig. 1a) before finally completing the usability questionnaire. In the laboratory study, the completion of the post-questionnaire was still interrupted by playing the video with the participant's verbalization. The attitude towards the catalog (3 items), the operator (3 items), the club vacation (3 items), the satisfaction with the catalog (5 items) and the intention to reuse the catalog and recommend it were operationalized as independent variables through one question each and tested.

Although both catalogs originate from club vacation operators in comparable price segments, they differ in both size and presentation. The Robinson catalog is a total of 251 pages in length, whereas the Aldiana catalog is 171 pages long. By way of example, important aspects in which both catalogs differ greatly and which could lead to different evaluations should be shown in the design: the Aldiana catalog works exclusively with black writing on white or light-colored backgrounds and also uses different colors within a category (not only across categories). Furthermore, it boasts an overall clear structure and the distinct differentiation of categories (e.g. club overviews) through the use of boxes. In addition, the Aldiana catalog contains several pictures per side (see fig. 2 Aldiana). In contrast, the color of the writing and background in the Robinson catalog is uniform (e.g. background light blue and writing dark blue), there are no rigid structures or boundaries in the overview pages and it focuses primarily on one picture per page (see fig. 2 Robinson).

3.4 Results: Usability scale

The results on the quality of the scale (CUQ) are presented below. In this, the reliability of the scale is examined as the initial quality criterion and the inter-correlations considered. The results are each compared to the existing findings of Konradt et al. (2003) from the extensive validation in an online context.

3.5 Reliability of the scale

All three sub-studies in the main study attest high to very high reliability (Crombach's a ≥ .928) for the scale used for usability using all 23 items. The reliabilities found are situated in the same area (Crombach's α ≥ .960) as the reliabilities found by Konradt et al. (2003) for the usability questionnaire in an online context (see table 2).

Table 3: Reliability of the overall scale usability main study

	Overall Scale Usability			
	Labortory study	Field study	Experts	Konradt et al. (2003)
Crombach's **α**	.928	.954	.958	.960

Furthermore, as in the case of Konradt et al. (2003), the inter-correlation of the scale was examined. The results found for the modified scale for usability used here also show that all scales correlate highly to extremely significantly with the overall scale (.51 ≤ r ≤ .98) and positively with each other (.34 ≤ r ≤ .85, see tables 3 and 4). The same pattern was also apparent in Konradt et al.'s original scale (see table 3) and speaks in favor of the quality of the modified scale.

Table 4: Inter-correlation of the overall usability scale with the single sub-dimensions

Overall Usability Value	Sub-values for Usability from the item categories (Dimensions)				
	1 General Usability Impression	2 Additional Information	3 Search	4 Club Overviews	5 Self-Description
Laboratory study	.96***	.58***	.74***	.85***	.59***
Field study	.97***	.58***	.66***	.94***	.66***
Experts	.98***	.79***	.81***	.89***	.51***
Konradt et al. (2003)	.92***	.62***	.84***	.71***	.79***

* p < .05¸ ** p < .01, *** p < .001

Table 5: Inter-correlation of the single sub-dimensions (Items) in the Laboratory (n1) and Field Study (n2)

	Sub-Values Usability from the Item Categories (Dimensions)			
	1 Gen. U. impression	2 Add. information	3 Search	4 Club overviews
	Lab | *Field*	Lab | *Field*	Lab | *Field*	Lab | *Field*
2 Add.Information	.60** | *.50****			
3 Search	.60*** | *.61****	.56** | *.34***		
4 Club overviews	.67*** | *.60****	.40** | *.40***	.57*** | *.59****	
5 Self-description	.57*** | *.85****	.56** | *.56****	.51*** | *.41***	.49*** | *.53****

* p < .05¸ ** p < .01, *** p < .001 ;

3.6 Construct validity

Construct validity signifies how capable an instrument is of illustrating the existing differences. These differences were determined on the basis of the independent expert survey prior to the actual main study with various instruments of usability evaluation. Consultants in travel agencies served as experts here. We can assume that they are capable of estimating how good a catalog is based upon their years of experience. Furthermore, the expert survey also represents a recognized instrument of measuring usability.

A total of 17 experts (travel agents, average experience *M* = 12.8 years) evaluated the two catalogs in the dimensions effectiveness, efficiency, satisfaction, overall usability and the five main-categories of the questionnaire used here independently of the main study. In the process, the Aldiana catalog was evaluated as user-friendly in all categories except "Additional Information on Travel Conditions". Both catalogs fall into the good to satisfactory categories and the experts' quantitative estimations lie close together. This raises the question as to whether the scale used here is capable of illustrating the differences in the two catalogs found by the experts in the preliminary study. This is clearly the case (cf. table 5). The Aldiana catalog is also evaluated to be considerably better than the Robinson catalog by the consumers in the sub-studies 1 and 2. Furthermore, the scale is capable of recording significant or potentially significant differences in the catalogs in all categories. Although the evaluations of the catalogs by the experts are similar in all categories in the preliminary study, the questionnaire is capable of differentiating between both catalogs (cf. table 5).

Significant differences could primarily be found in the areas in which the catalogs differ greatly due to the design, the club overviews and the possibility of finding information quickly (cf. table 5).

The catalog which emerged from this study as user-friendly was evaluated again by experts within the context of a third study. This served the repeated examination of the results obtained by experts. This time, the complete usability scale was used (identical to that of sub-studies one and two). The results obtained support both the estimations of the consumers and the global estimations of the pre-study.

Table 6: Significant differences between the catalogs found by the CUQ; samples n1 and n2

	Significant differences between the catalogs	
	df, T	**p**
Usability Overall Value (all 23 Items)	130, 2.38	.019
General Usability Impression (13 Items)	139, 1.82	.070
This catalog makes searching for a club vacation / hotel easy	148, 1.96	.049
You can quickly gain an overview of the catalog	148, 1.87	.063
The catalog is well structured	147, 1.83	.070
I can always recognize important information quickly	147, 1.92	.063
I can always find information which is important to me quickly	148, 2.49	.014
Additional Information on the Travel Conditions (1 Item)	142, 1.82	.071
Search for Clubs / Hotels (1 Item)	147, 1.70	.062
Club / Hotel Overviews (7 Items)	137, 2.75	.007
The club / hotel overview pages are clear	146, 2.91	.004
The club / hotel overview categories offered are selected sensibly	146, 1.67	.098
The symbols on the club / hotel overview pages are comprehensible	148, 2.59	.011
I always have an overview about the price of the club / hotel	143, 2.95	.004
I find the pictures used helpful	147, 1.87	.063
Self-description (1 Item)	148, 1.68	.095

On the basis of verbal protocols, the amount of the usability expressed for each catalog was determined for each participant. Furthermore, the positive statements on usability were also counted. As we can gather from table 7, the overall usability value of the catalogs (recorded via the questionnaire) correlates positively with the number of usability problems expressed from the verbal protocol (r = .489). This shows that the participants' usability problems identified in the questionnaire are also to be found in the verbal protocols. The participants who assessed the usability as worse also expressed more problems in the verbal protocols. Furthermore, the number of positive statements correlates negatively with the overall usability value, as was to be expected, but the correlation is not significant.

Table 7: Correlation between the overall usability value and the number of the negative and positive statements on usability on the basis of verbal protocols (n= 92)

	Negative Statement on Usability (Usability Problems)		Positive Statements on Usability	
	r	p	r	p
Overall Usability Value	.489***	< .001	-.153 *n.s.*	.184

n.s. not significant, *** p < .001 ;

3.7 Criteria validity

Subsequently, the results of the correlation analysis between the usability and effects are presented. These effects, like reuse and attitudes, are classed as an indicator for the criteria validity. The two intentions and the attitudes were selected as criteria, because other studies proved their correlations with usability (Lee and Lee 2004; Kohavi and Round 2004; Büttner, Schulz, and Silberer 2006). Furthermore, users' satisfaction with the catalog usage was considered because it is a partial aspect of usability according to the ISO norm and, consequently, strong correlations are indicators of scale's criteria validity (Schweibenz & Thissen 2003).

Table 8: Correlations of the usability with effects of the usability

Construct / Item	Usability		
	Laboratory study	Field study	Experts
	r	r	r
Reuse	.477***	.576***	n/a[a)]
Recommendation	.546***	.617***	n/a
Satisfaction	.697***	.893***	.721**
Attitude to catalog	.665***	.842***	.849***
Attitude to operator	.331**	.530***	.497*

* p < .05, ** p < .01, *** p < .001; [a)] Variable not surveyed in the study

The results support the validity of the scale. As we can see in table 8, all the correlations found are highly to very highly significant and are situated on a medium to high level. With r ≥ .697, the satisfaction correlates highly significantly with the usability, which was to be expected on the basis of the definitional proximity (cf. ISO 9241). Furthermore, the correlation between both attitude factors and the usability is significant. In this, like in a study on trust, risk and usability by Büttner, et al. (2006), the attitude towards the catalog correlated highly significantly with the usability at a high level with r ≥ .665. The expected influence of the usability on the attitude towards the operator can also be confirmed, whereby this is less striking. If we assume that the catalog is the first and only source of information on the operator for many participants, the results obtained are explicable. In addition, the fact that a significant correlation between intention and usability could be found in the studies by Lee and Lee (2004), Konradt et al. (2003) and the study by Büttner et al. (2006) in an online context and can also be confirmed here in the context of print catalogs supports the validity of the scale (see table 5).

3.8 Results: Verbal protocols

Besides the measurement on the basis of the questionnaire, we will briefly address the results which were obtained with the aid of the verbal protocols. A total of 380 expressed usability problems were recorded in the context of the coding of the verbal protocols of the 92 participants. On average, every participant thus expressed *M*= 4.30 (*SD*= 2.98) usability problems.

However, how the number of expressed usability problems is distributed across the categories in (see category system 3.2) is interesting (see table 9). Category 1 *impression and validity of the pictures* occupies first place with 113 negative statements (29.7%) and thus appears to have the highest proportion of a negative usability assessment. The high number of namings also indicates a significant need for action in the corresponding category. Category 4, *information content of the texts* (25.0%), and category 5, *overall composition of the catalog* (23.9%), subsequently follow.

Similarly, the previous results can also be measured using the questionnaire scale. At this point, the verbal protocols nevertheless offer advanced possibilities for analysis. It is possible to address relevant categories in more detail through a qualitative evaluation of the statements. Consequently, the problems can be identified concretely and changes implemented more easily on the basis of the statements. Not only do we know *that there is a problem* with the pictures; we also know the reason for the problem. Corresponding statements of the participants are e.g. “the pictures only show small sections of the beach...“, “the resort has been photographed from unfavorable positions...“ or “the photos seem old...“.

Table 9: Number of Negative Statements on Usability

Category	Percent % (Number of namings)
1: Impression and expressiveness of the pictures & photos	29.7 (113)
2: Clarity and design of the colors	18.7 (71)
3: General color scheme	2.6 (10)
4: Information content of the texts	25.0 (95)
5: Overall composition of the catalog	23.9 (91)

Finally, it can be noted that the conclusions and evaluation possibilities of the verbal statements are very big. Above all, a qualitative analysis of the statements or a process-related analysis (e.g. sequence analysis) of the interaction process promises substantial information.

4. Conclusion

Both the reliability and the validity support the scale developed here to measure the usability of print catalogs. In the process, the results on reliability coincide with those of the previous studies (Konradt et al. 2003). Furthermore, the scale succeeds in illustrating the expected differences in the main study on the basis of an expert survey and analysis of the design. The results based on the verbal protocols also coincide with measurements of the usability recorded through the questionnaire. In addition, the results of the correlation analysis between the overall value of the usability and effects also coincide with known results from the field of usability research. The highly significant correlations found between the intentions and the usability are primarily interesting for industry and especially marketing management. Consequently, this result underlines the importance of a user-friendly catalog design and the necessity for testing. After all, a user-friendly catalog encourages the customer to recommend it, reuse it and ultimately book a vacation. The use of verbal protocols facilitates a deeper insight into the actual interaction process of the user. So it can help identify more detailed usability problems than an overall evaluation via a usability scale. However, the effort required for the data coding is above all immense.

5. Implications for the process of catalog development and control

The aim of the study presented here was the development of an inexpensive and handy instrument to test the usability of print catalogs. With this scale, it is possible to evaluate catalogs in a standardized process and to get useful information on problem areas with quite low effort. Consequently, it is possible to reveal reliable and valid differences of catalogs and assess their quality. The scale offers the possibility of benchmarking different catalogs, a rapid comparison of different prototypes or also the possibility of preparing for advanced and final user tests. Furthermore, the evaluation by a large number of users without too much effort is also conceivable. One possibility of implementing this would be the online presentation of the catalog pages with a corresponding link to the questionnaire. After consideration of the pages, the corresponding participant's data would be allocated directly to a data bank for evaluation.

Further experiences with the scale would be preferable and necessary for the future. It's possible to transfer the instrument to other catalog forms. However, a modification or adaptation of different items might be necessary depending on the aim of the catalog. The use of the scale in the evaluation of other communication instruments, such as personalized letters or instructions for use, would be equally as conceivable.

Limits to the interpretability of the usability problems revealed are also set through the given dimensions of the scale. Concrete insights into the behavior and experience in the actual interaction with the print catalog cannot be obtained in this manner. In each case, it is a question of global judgment on the overall use process. If concrete insights are required, the use of verbal protocols can provide a more detailed view. As we have seen, the method provides the opportunity to record the behavior and experience accompanying the process and gain further insights. However, the effort which the verbal protocols require should not be underestimated. A use in a small part of the overall sample seems sensible and should provide sufficient information. A combination of the two methods consequently appears useful.

Therefore, the scale developed represents a useful addition to qualitative methods. In this way, it should be possible to design print catalogs from different industries in an even more user-friendly manner in future and consequently attract customers.

References

Büttner, O. B., Schulz, S. and Silberer, G. (2006) "Perceived Risk and Deliberation in Retailer Choice: Consumer Behavior towards Online Pharmacies", *Advances in Consumer Research*, Vol. 33, ed. by C. Pechmann and L. Price, Duluth, MN: Association for Consumer Research, pp197-202.

Büttner, O. B. and Silberer, G. (2007) *Video-cued Thought Protocols – A Method for Tracing Cognitive Processes at the Point of Purchase*, in: Lowrey, T. M. (Ed.), Brick & Mortar Shopping in the 21st Century, Lawrence Erlbaum Associates, Mahwah, NJ

Bhv (2005) „Pressemitteilungen des Bundesverbandes des Deutschen Versandhandels e.V. date: 22.11.2005 & 9.12.2005"; [online], http://www.versandhandel.org/Presse.89.0.html

ISO (1998) "ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability", [online], http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16883&ICS1=13&ICS2=180&ICS3=

Ericsson, K. A. and Simon, H. A. (1993) *Protocol Analysis – Verbal Reports as Data*, Cambridge, The MIT Press

Fodness, D. and Murray, B. (1999) “A model of tourist information search behaviour”, *Journal of Travel Research*, Vol. 37, No. 1, pp220-230
Goossens, C. (1994) “External information search: Effects of tour brochures with experiential information”, *Journal of Travel and Tourism Marketing*, Vol. 3, No. 3, pp89-107
Hodgson, P. (1993) “Tour operator brochure design research revisited”, *Journal of Travel Research*, Vol. 32, No. 1, pp50-52
Hodgson, P. (1991) “Qualitive research on tour brochure design”, Journal of Market Research Society, Vol. 33, No. 1, pp51-55
Kirakowski, J. (1996) *The software usability measurement inventory*, in: Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McCelland, I. L. (Ed.). Usability evaluation in Industry, Taylor & Francis, London, pp169-177
Kohavi, R and Round, M. (2004) “Emetrics Summit 2004: Front Line Internet Analyticsat Amazon.com”, [online], http://ai.stanford.edu/~ronnyk/emetricsAmazon.pdf
Konradt, U., Wandke, H., Balazs, B. and Christophersen, T. (2003) “Usability in online shops: scale construction, validation and influence on the buyers intention and decision”, *Behaviour & Information Technology*, Vol. 22, No. 3, pp61-73
Krömker, H. (1999) *Die Welt der Benutzerfreundlichkeit*, in: Henning, Jörg & Tjarks-Sobhani, Marita (Ed.). Verständlichkeit und Nutzungsfreundlichkeit von technischer Dokumentation, Lübeck, Schmidt-Römhild, pp22-33
Lee, B.-K. and Lee, W.N. (2004) “The effect of information overloadon Consumer Choice Quality in an On-Line Environment”, *Psychology & Marketing*, Vol. 21, No. 3, pp159-183
Mack, R. L. and Nielsen J. (1994) *Executive Summary*, in: Nielsen, J. & Mack, R. L. (Ed.). Usability Inspection Methods, Wiley, New York, pp1-23
Nielsen, J. (1994) “Guerilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier”, [online], http://www.useit.com/papers/guerrilla_hci.html
Nielsen, J. (1993) *Usability Engineering*, Academic Press, Bosten, San Diego, New York etc.
Reynolds, F. D. (1974) “An Analysis of Catalog Behavior”, *Journal of Marketing*, Vol. 38, No. 3, pp47-51
Schmeißer, D. R., Behrendt, C. and Singer, B. (2005) „Werbewirkung messen und optimieren – Wegweiser und Methoden am Beispiel von Print-Anzeigen“, *Planung & Analyse*, No. 6/2005, pp49-56
Schweibenz, W. and Thissen, F. (2003) *Qualität im Web – Benutzerfreundliche Webseiten durch Usability Evaluation*, Berlin, Heidelberg, New York, Springer Verlag
Silberer, G. (2006) “A New Approach to Assessing Cognitive Processes at the Point of Sale – Video-Cued Thought Protocols”, *European Advances in Consumer Research*, Vol. 7, ed. By K. M. Ekström and H. Brembeck, pp555-556
Stoessel, S. (2002) *Methoden des Testings im Usability Engineering*, in: Beier, Markus (Ed.). Usability: Nutzerfreundliches Web-Design, Springer, Berlin, pp75-96
Sweeney, M., Maguire, M. and Shackel, B. (1993) “Evaluating User-Computer interaction: A Framework”, *International Journal of Market Man-Machine*, Vol. 38, No. 4, pp689-711
Yom, M. and Wilhelm, T. H. (2004) *Methoden und Erkenntnisse der Web-Usability-Forschung*, in: Zerfaß, A. & Zimmermann, H. (Ed.). Usability von Internet-Angeboten. Grundlagen und Fallstudien, Baden-Würtenberg, Medien und Filmgesellschaft

