

Assessing Qualitative Usability in life-threatening, time-critical and unstable Situations

Simon Nestler* Eva Artinger* Tayfur Coskun* Yeliz Yildirim-Krannig‡
Sandy Schumann‡ Mareike Maehler‡ Fabian Wucholt‡
Stefan Strohschneider‡ Gudrun Klinker*

*Fachgebiet Augmented Reality ‡Professur für Interkulturelle Kommunikation
Technische Universität München Friedrich-Schiller-Universität Jena
Fakultät für Informatik Interkulturelle Wirtschaftskommunikation
Boltzmannstraße 3 Ernst-Abbe-Platz 8
85748 Garching bei München 07743 Jena
nestler|klinker@in.tum.de stefan.strohschneider@uni-jena.de

Abstract:

We developed a heuristic for assessing the usability of mobile user-interfaces in life-threatening, time-critical and unstable situations in a qualitative way. The major advantages of our approach as opposed to standardized quantitative questionnaires is the independence from a baseline, the possibility to make absolute statements and the potential for adaptations. When creating a qualitative semi structured interview we adhere to the common modus operandi of the qualitative social research. On the basis of 17 common quantitative questionnaires on usability we identified the five major categories **Utility**, **Intuitiveness**, **Memorability**, **Learnability** and **Personal Effect**. We selected all questions from the questionnaires which are useful for assessing the usability of user-interfaces in emergencies. Furthermore, we rephrased the closed-ended questions to open-ended ones. The quantification of research results is possible by weighting the qualitative results in dependence to the research question.

1 Introduction

Within the scope of the *SpeedUp* project¹ we found out that the sound evaluation of mobile user-interfaces for medical emergencies is challenging for three reasons: (1) Mobile user-interfaces replace paper based workflows, (2) evaluations take place in lifelike trainings and (3) stress is dominant in medical emergencies.

Paper based workflows. When mobile user-interfaces are compared to paper based approaches, this comparison is inhomogeneous. Although mobile user-interfaces increase

¹The project SpeedUp is funded by the German Federal Ministry of Education and Research (BMBF) within the programme "Research for Civil Security" (May 1st, 2009 - April 30th, 2012, FKZ: 13N10175). Website: <http://www.speedup-projekt.de>

the quality of information², entering high quality information is more laborious. By comparing the usability of the mobile user-interface to the usability of paper, the mobile user-interface is in an inferior position. It is essential to provide a possibility to evaluate mobile user-interfaces without needing a baseline (*see A*).

Real life scenarios. Due to the fact that evaluations of mobile user-interfaces for medical emergencies take place in lifelike trainings, the repeatability of the evaluation is limited. The trainer cannot completely control the set of parameters in these scenarios. Consequently the sequential, quantitative comparison of different design alternatives is subject to restrictions. Evaluating user-interfaces without requiring a quantitative comparison of different alternatives is essential (*see B*).

Dominance of stress The impact of stress on the usability is not considered by the different standardized questionnaires. Questionnaires either focus on usability or on physical and mental demands. However, because usability depends on the task load and the mental demands are high in medical emergencies, considering the impact of stress is essential for gaining meaningful results. Furthermore, weighting the different categories is essential for evaluating mobile-user interfaces (*see C*).

Consequently for the proper evaluation of user-interfaces in life-threatening, time-critical and unstable situations a new type of usability evaluation is required. We propose the conduction of a qualitative semi structured interview for three reasons:

- qualitative interviews **do not depend on a baseline**³ (A)
- qualitative interviews **do not require different alternatives** (B)
- qualitative interviews **allow the weighting of categories** (C)

Furthermore, qualitative information is essential for improving the capabilities of mobile user-interfaces. By providing detailed qualitative information, the engineers and computer scientists can identify the weaknesses of the mobile user-interfaces more easily and can improve the mobile user-interfaces more effectively and efficiently. Consequently the application of qualitative assessments in requirements analysis, interaction design, prototypical implementation and evaluation simplifies the overall process of developing mobile user-interfaces.

2 Related work

In literature different standardized questionnaires are used for the evaluation of user-interfaces. In the following we present the 17 most common quantitative questionnaires with focus on *Usability*, *Attractiveness*, *Satisfaction*, *Experience* and *Work Load*. Some of these questionnaires categorize the different questions. All categories which are transferred to our qualitative semi structured interview are written in bold.

²Quality of information is enhanced by increased structuring

³the qualitative properties of user-interfaces are absolute

The *After-Scenario Questionnaire* consists of three questions on the user's **Satisfaction** [Lew91]. The *AttrakDiff* questionnaire consists of 21 pairs of antithetic adjectives. The *AttrakDiff* focuses on **Attractiveness**, *Hedonic Quality* and *Pragmatic Quality* [HBK03]. The *Computer Literacy Scale* consists of different questions on the user's experience with computers. The *CLS* focuses on *Experience*, *Symbols* and *Terminology* [SD08]. The *Computer System Usability Questionnaire* consists of 19 questions on the system's usability. The *CSUQ* is unstructured and does not use categories [Lew95]. The *IsoMetrics* focuses on usability in general. The *IsoMetrics* focuses on *Adequacy of Tasks*, *Ability of Self-Characterization*, *Controllability*, **Compliance with Expectations**, **Error Robustness**, **Customizability** and **Learnability** [GH99, GHD99]. The structure of *Isonorm 9241-10* is quite similar to *IsoMetrics*. The categories are identical to *IsoMetrics* – except *Fault Tolerance* (instead of **Error Robustness**) and *Ease of Learning* (instead of **Learnability**) [Prü97]. The *Nielsen's Attributes of Usability* consist of 5 different categories: **Learnability**, **Efficiency**, **Memorability**, **Errors** and **Subjective Satisfaction** [Nie93]. The *Nielsen's Heuristic Evaluation* consists of 10 questions which result in a heuristic guideline. The *NHE* is unstructured and does not use categories [NM90, Nie94a, Nie94b]. The *Practical Heuristics for Usability Evaluation* consist of a heuristic guideline with 13 questions. The *PHUE* focus on **Learning**, *Adapting to the user*, **Feedback** and **Errors** [Per94]. The *Perceived Usefulness and Ease of Use* questionnaire consists of 12 questions on *Usefulness* and **Ease of Use** [Dav89]. The *Purdue Usability Testing Questionnaire* consists of 100 questions in different categories. The *PUTQ* focuses on *Compatibility*, **Consistency**, **Flexibility**, **Learnability**, *Minimal Action*, *Minimal Memory Load*, *Perceptual Limitation* and **User Guidance** [LCS97]. The *Questionnaire for User Interface Satisfaction* consists of 27 questions on **Satisfaction**. The *QUIS* focuses on *Overall Reaction*, *Screen*, *Terminology*, *System Information*, **Learning** and *System Capabilities* [CDN88]. The *Software Usability Measurement Inventory* consists of 50 questions on usability in general. The *SUMI* is unstructured and does not use categories [Kir96]. The *System Usability Scale* consists of 10 questions on usability. The *SUS* is unstructured and does not use categories [Bro96]. The *NASA Task Load Index* consists of 6 questions on work load. The *NASA-TLX* is unstructured and does not use categories [HS88]. The *User Experience Questionnaire* consists of 26 pairs of antithetic adjectives. The *UEQ* focuses on **Attractiveness**, **Perspicuity**, **Novelty**, *Stimulation* and **Dependability** [LHS08]. The *USE Questionnaire* consists of 30 questions on general usability. The *USEQ* focuses on *Usefulness*, **Ease of Use**, **Ease of Learning** and **Satisfaction** [Lun01].

When taking a closer look on the categories from these questionnaires it becomes clear, that these various questionnaires are not selective. Several aspects, however, such as **Stress**, **Expericence** or **User Guidance**, are only considered by one questionnaire.

3 Method

For the qualitative evaluation of user-interfaces we make use of a method which comes from social science: qualitative interviews. According to [KDSR07] these qualitative interviews take the perspective of the subjects into consideration. Besides the parameters

the underlying causes of decisions are subject of the evaluation. The used methods are legitimated by their contribution to the solution of a research question. The semi-structured interviews base on a set of problems, each of these problems consists of a set of questions. These questions are orally answered by the subject and the interview is documented on a voice recorder. The interviewer uses open-ended questions and avoids interrupting the subject. The aim of the interview is to discuss all problems with the subject⁴.

Although this method is quite well known and successful in social sciences, it has not found it's way into usability research. Therefore, we transferred this method from social sciences to usability research. Our qualitative interview on usability was developed on the basis of an extensive brainstorming process in combination with the creating of a detailed associagram. During the brainstorming we made use of the quantitative questionnaires on usability. Furthermore, the categories of the quantitative questionnaires was utilized in our associagram. Due to the fact that we started from the scratch we could design a method which is independent from a baseline (*see A*) and does not rely on different alternatives (*see B*). The adaption of the questionnaire to various fields of application can be guaranteed by a flexible weighting of the different categories (*see C*). Consequently this method can flexibly be adapted to the special requirements on mobile user-interfaces for emergencies.

Qualitative evaluation is not limited to an effectivity control, assessing the value of components or the overall object is of equal importance. The evaluation has to conform with critical-rational demands as well as with ethic-moral standards. In general an evaluation can have four different aims [Kin98]: (R.1) facilitating insights, (R.2) reaching decisions, (R.3) legitimating decisions and (R.4) optimizing objects. In usability engineering the aspects (R.1) and (R.4) are of special importance during the prototyping phase.

[KDSR07] describes the general principles for preparing qualitative interviews: (1) The research question has to be concretized, (2) the questions have to be selected by a team of experts, (3) questions have to be formulated open-ended. Furthermore, the evaluation design has to consider the following questions: (D.1) How many subjects should be interviewed, (D.2) how are these subjects chosen, (D.3) when should the evaluation take place, (D.4) how are the interviews recorded and (D.5) how will the transliteration done. The most popular literature on the number of subjects was published by [NL93], [Vir92] and [Lew94]. In these publications a mathematical model is developed which enables the user-interface designer to calculate the optimal sample size: $U = 1 - (1 - p)^n$. The probability to detect a problem (U) depends on the probability (p) that a subject identifies a problem and the number (n) of subjects. Due to the fact that the probability (p) is not known for qualitative usability evaluations, we have to rely on their general heuristics: (1) most usability problems are detected with three to five subjects, (2) it is unlikely that additional subjects reveal new information, (3) most severe usability problems are detected by the first few subjects. Consequently we use three to five subjects for the qualitative usability evaluations (D.1). We choose this set of subjects randomly from the group of our end users (D.2). The evaluations take place in each iteration⁵ (D.3). The interviews are documented by voice recorders (D.4) and are manually transliterated (D.5).

After the transliteration of the interviews, the different statements are categorized accord-

⁴Usually it is not necessary to ask the complete set of questions to cover all problems

⁵An iteration consists of requirements analysis, interaction design, prototypic implementation and evaluation

ing to the following rules: (1) categories are terms, (2) categories are deduced from the aims and research questions, (3) categories should neither be slender nor extensive and (4) categories have to be selective [May08]. According to [Kin98] the following requirements have to be taken into consideration in order to get methodically dependable and valid evaluation results: (1) The individual cases are part of the research process, (2) research process is open for revision and extension, (3) the general procedure is lead by a set rules, (4) research processes are seen as an interaction, (5) the objects are analyzed holistically and (6) generalization is demonstrated by arguments. Consequently the presented set of problems and questions is always subject of further research.

4 Results

The combination of existing quantitative questionnaires on usability and qualitative research methods leads to semi structured interviews on usability. In the following the resulting categories and questions of the qualitative interview are described. The process of generating categories and open-ended questions was performed according to the process from [KDSR07] as presented above.

4.1 Categories

On the basis of these questionnaires three usability experts conjointly identified five main categories for the qualitative interview: **(U) Utility**, **(J) Intuitiveness**, **(M) Memorability** [Nie93], **(L) Learnability** [GH99, GHD99, Prü97, Nie93, Per94, LCS97, CDN88, Lun01] and **(P) Personal Effect**. These main five categories are subdivided in four to five sub-categories each: The **(U) Utility** is classified in: (U-I) Dependability [LHS08, GH99, GHD99, Prü97], (U-II) Ease of Use [Dav89, Lun01], (U-III) Effectiveness, (U-IV) Efficiency [Nie93] and (U-V) Productivity. The **(J) Intuitiveness** is classified in: (J-I) Affordance [GH99, GHD99, Prü97], (J-II) Expectations [GH99, GHD99, Prü97], (J-III) Conventions and (J-IV) Transparency. The **(M) Memorability** is classified in: (M-I) Consistency [LCS97], (M-II) Customizability [GH99, GHD99, Prü97, LCS97], (M-III) Complexity and (M-IV) Perspicuity [LHS08]. The **(L) Learnability** is classified in: (L-I) Error handling [GH99, GHD99, Prü97, Nie93, Per94], (L-II) Feedback [Per94], (L-III) Help and (L-IV) User Guidance [LCS97]. The **(P) Personal Effect** is classified in: (P-I) Attractiveness [HBK03, LHS08], (P-II) Novelty [LHS08], (P-III) Satisfaction [Lew91, Nie93, CDN88, Lun01] and (P-IV) Stress.

These categories fulfill the major requirements from [KDSR07]: The categories are terms, are deduced from the research question and are selective. These terms are directly taken from the different questionnaires. Due to the fact that these usability questionnaires deal with our research question, we could prove that the categories are deduced from the research question. Furthermore, the concordant identification of the five main categories by three usability experts indicates the selectiveness of these categories. The question whether

these categories are neither slender nor extensive has to be proven within the scope of an evaluation. Table 1 gives an overview on all categories and sub-categories.

4.2 Questions

In the next step the three usability experts conjointly assigned all questions from the qualitative usability questionnaires to the different categories. When the assignment was ambiguous⁶ the question was removed from the qualitative semi structured interview.

The resulting semi structured interview is shown in the appendix. We followed the principles from [KDSR07]: The concrete research question is "usability", the questions were selected by a team of three experts with regard to the research question and finally the closed-ended questions were rephrased to open-ended questions. This rephrasing is necessary to be able to use these questions as a guideline for the semi-structured interview. In each of the categories several different questions are available and the interviewer is free to choose a subset of questions which fits best for the concrete user-interface. As an alternative the team of interviewers can pre-select a subset of questions in the run-up to the evaluation.

5 Discussion

The qualitative evaluation provides a detailed assessment of the quality of a mobile user-interface. According to [May93] a quantification of research results is important. A quantitative score, however, is not directly deduced by the proposed method. [Kin98] proposes a quantitative analysis of the qualitative evaluation to receive the required quantitative data. In this analysis the transliterated statements from the interview are sorted by experts into the categories and sub-categories. This quantitative summarization of the qualitative evaluation is the basis for the quantification of the research results according to [May93]. The summarized data is adjusted on a 3-point scale: (a) positive comment (1.0), (b) neutral comment (0.5) and (c) negative comment (0.0). The mean value over all statements in the same sub-categories is calculated. As a result we receive a quantitative rating of all sub-categories on a scale from 0.0 to 1.0.

After calculating these scales, an usability scale which is application specific can be calculated by weighting these categories. The categories for the quantitative summarization are shown in Table 1. The balanced weighting of the categories and sub-categories from Table 1 is shown in Table 2. When performing a general evaluation of user-interfaces this weighting leads to a quantitative one-dimensional usability value – besides the qualitative results.

Due to the dominance of stress in emergencies we argued that weighting the different sub-categories is essential for evaluating mobile-user interfaces. These two aspects can be

⁶An unambiguous assignment requires at least the same categorization by two of the three experts

Utility	Intuitiveness	Memorability	Learnability	Personal Effect
Dependability	Affordance	Consistency	Error handling	Attractiveness
Ease of Use	Expectations	Customizability	Feedback	Novelty
Effectiveness	Conventions	Complexity	Help	Satisfaction
Efficiency	Transparency	Perspicuity	User Guidance	Stress
Productivity				

Table 1: Categories for Quantitative Summarization

20 ^a	20 ^b	20 ^c	20 ^d	20 ^e
4	5	5	5	5
4	5	5	5	5
4	5	5	5	5
4	5	5	5	5
4				

Table 2: Balanced weighting (general user-interfaces)

^aUtility
^bIntuitiveness
^cMemorability
^dLearnability
^ePersonal Effect

35 ^a	10 ^b	15 ^c	10 ^d	30 ^e
7	3	5	3	2
7	3	0	3	2
7	2	5	2	6
7	2	5	2	20
7				

Table 3: Focused weighting (user-interfaces for emergencies)

^aUtility
^bIntuitiveness
^cMemorability
^dLearnability
^ePersonal Effect

brought together by weighting the qualitative results in dependence to the research question. For the evaluation of mobile user-interfaces for emergencies we changed the weighting with regard to the research question as shown in Table 3. The **Utility** and the **Personal Effect** is of special importance in emergencies [Nes10]. Therefore, the weights for these categories were increased. **Stress** is dominant in emergencies, whereas **Attractiveness** and **Novelty** is of lower importance. Therefore, the weight for **Stress** was increased as well. From previous requirements analyses we know that **Customization** is difficult in emergencies [Nes08, Nes10], consequently it is not considered in the emergency specific usability value. Finally, the usability score is calculated by multiplying all weights with the quantitative scores of the sub-categories:

$$\begin{aligned}
 U &= \sum_{c \in C} \left(U_c * \frac{w_c}{100} \right) = \sum_{c \in C} \left(U_c * \sum_{s' \in S(c)} \frac{w(s')}{100} \right) = \\
 &= \sum_{c \in C} \left(\sum_{s \in S(c)} \frac{w(s) * 100}{\sum_{s'' \in S(c)} w(s'')} * v(s) * \sum_{s' \in S(c)} \frac{w(s')}{100} \right) = \sum_{c \in C} \left(\sum_{s \in S(c)} w(s) * v(s) \right) = \sum_{s \in S} (U_s * w_s)
 \end{aligned}$$

Besides the general usability score U , more specific scores U_c for each category c can be calculated – as described above. For the calculation of these specific scores the weights have to be normalized. The score U as well as the scores U_c go from 0 to 100, due to the fact that the sum of all weights is 100.

6 Conclusion and future work

The qualitative usability evaluation leads to a benefit for developers, because detailed qualitative information is provided. This qualitative information helps to identify the weaknesses of the mobile user-interfaces more easily. Nevertheless a quantitative usability score can be provided as well and the qualitative usability evaluation has no disadvantages – but many benefits – as opposed to the quantitative usability evaluation. Furthermore, the quantitative score can flexibly be adapted to the concrete research question⁷.

We will use our qualitative interviews to get deeper insights in the human-computer interaction in emergencies (R.1, see above). Furthermore, the existing research method is iteratively improved by all these qualitative evaluations due to the fact that a re-categorization of the answers is performed subsequent to every interview. The questionnaire can be simplified by reducing the number of questions and can be customized by weighting the different categories. Due to the fact that a small set of subjects (three to five) is used in qualitative evaluations, the evaluation efforts are reduced significantly.

In the future we expect an intensive use of qualitative usability evaluations in the ubiquitous computing domain because of the following reasons:

- ubiquitous applications **are new and innovative**
- ubiquitous applications **are frequently evaluated in real-life scenarios**
- ubiquitous applications **are developed domain-specific**

When building new and innovative ubiquitous applications, the comparison with existing applications is often difficult. On the one hand innovative applications exceed the capabilities and functionalities of existing ones and on the other hand subjects are more familiar with existing applications. Consequently these qualitative, comparative evaluations are often inhomogeneous in the ubiquitous computing domain. In our impression qualitative assessment of the attributes is more promising with regard to the effective improvement of the ubiquitous application. When evaluating in real-life or lifelike scenarios, the comparability of successive runs is limited. When the ubiquitous application focuses on a specific domain, the usage of standardized questionnaires is complicated. Consequently the flexible customizability of the qualitative evaluation is a strong argument for its future, more intense application.

7 Appendix: Questionnaire

Id	Qualitative statement	Reference
U	Utility	
U.1	How do you use the software in your job?	[Dav89]
U.2	What is the use of the software?	[Lun01]

⁷In our case: mobile user-interfaces for emergencies

Id	Qualitative statement	Reference
U-I	Dependability	
U-I.1	What can you say about your confidence in the software?	[Bro96]
U-I.2	What do you think about the software's security?	[LHS08]
U-I.3	What do you think about the software's reliability?	[CDN88]
U-I.4	Who initiates the sequence control?	[LCS97]
U-I.5	Who is in command of this software?	[Kir96]
U-II	Ease of Use	
U-II.1	How do you reach your goals with the software?	[Kir96]
U-II.2	How can you shift among different windows?	[LCS97, GH99]
U-II.3	How can you change between different parts of your tasks?	[Kir96, GH99, Prü97]
U-II.4	How do you find required information with this software?	[Lew95]
U-II.5	What kind of help did you need before using the software?	[Bro96]
U-II.6	What do you think about the understandability of information?	[Lew95]
U-II.7	Which parts of the software are easy to use?	[Dav89, Bro96, Lun01]
U-II.8	Which parts of the software are cumbersome to use?	[Bro96, CDN88]
U-II.9	Why is the use of this software simple?	[Lun01]
U-III	Effectiveness	
U-III.1	How can you effectively work with this software?	[Lew95]
U-III.2	How helps the information you in completing your tasks?	[Lew95]
U-III.3	To what extend enhances the software your effectiveness?	[Dav89]
U-III.4	What is the software's influence on your effectiveness?	[Lun01]
U-IV	Efficiency	
U-IV.1	How hard did you struggle to accomplish your aims?	[Nie93]
U-IV.2	How many steps do you need to accomplish your aims?	[HS88]
U-IV.3	How cumbersome is the cursor positioning?	[Lun01, GH99]
U-IV.4	How is the minimization of steps in menus done?	[LCS97]
U-IV.5	How extensive are the user control actions?	[LCS97]
U-IV.6	How are frequent control entries accessed?	[LCS97, GH99]
U-IV.7	How can the user return to higher-level menus?	[LCS97]
U-IV.8	To what extend unnecessary inputs have to entered?	[Prü97, LCS97]
U-IV.9	To what extend does the software interrupt your tasks?	[Prü97]
U-IV.10	To what extend is the software economic with keystrokes?	[Kir96]
U-IV.11	What do you think about the response to your inputs?	[Kir96]
U-IV.12	What do you think about the software's speed?	[Kir96, CDN88, Dav89]
U-IV.13	What kind of shortcuts are provided?	[NM90]
U-IV.14	What influence has the software on process times?	[Lun01, LHS08]
U-IV.15	What kind of default values are available?	[LCS97]
U-IV.15	What kind of searches are available?	[LCS97]
U-IV.16	What can you say about the efficiency of this software?	[LHS08, Lew95, Prü97]
U-V	Productivity	
U-V.1	How is the accomplishment of your tasks supported?	[Lun01]
U-V.2	How long did it take until you became productive?	[Lew95]
U-V.3	How do you get something to work?	[Kir96]
U-V.4	How are connected work packages handled?	[GH99]
U-V.5	How does the visualization support the handling of your tasks?	[GH99]
U-V.6	How can this software be used in an optimal way?	[GH99]
U-V.7	How successful were you in accomplishing your tasks?	[HS88, Lun01]
U-V.8	To what extend is the software practical?	[HBK03]
U-V.9	To what extend is the solution of your tasks supported?	[GH99]
U-V.10	What is the software's influence on your productivity?	[Lun01, Dav89]
J	Intuitiveness	
J-I	Affordance	
J-I.1	How understandable is the wording in the software?	[GH99]
J-I.2	How does the software indicate which inputs are possible?	[Prü97]
J-I.3	How dependent are you on written instructions?	[Lun01]
J-I.4	To what extend are the command names meaningful?	[LCS97]

Id	Qualitative statement	Reference
J-I.5	To what extend is this deactivation of elements clearly visible?	[GH99]
J-I.6	To what extend can software used without documentation?	[Per94]
J-I.7	What kind of affordances are provided by the software?	[Per94]
J-I.8	What do you think about the mapping of functionalities?	[Per94]
J-I.9	Which terms, abbreviations or symbols are complicated?	[Prü97]
J-II	Expectations	
J-II.1	How easy can you get things done which you want to do?	[Dav89, Lun01]
J-II.2	To what extend does the software meet your expectations?	[LHS08, GH99]
J-II.3	To what extend is the software predictable?	[HBK03, GH99]
J-II.4	To what extend are the process times predictable?	[GH99, Prü97]
J-II.5	What is the relation between expectations and behavior?	[LCS97, Lun01]
J-II.6	When did the software not do what you were expecting?	[Kir96]
J-II.7	When did the software stop unexpectedly?	[Kir96]
J-II.8	Which functions and capabilities do you expect?	[Lew95]
J-III	Conventions	
J-III.1	How is the menu selection done?	[LCS97]
J-III.2	How are abbreviations and acronyms used?	[LCS97]
J-III.3	How is the consistency with user conventions guaranteed?	[LCS97]
J-III.4	To what extend is the software compatible with conventions?	[LCS97, LHS08, HBK03]
J-III.5	To what extend are the color codes conventional?	[LCS97]
J-III.6	To what extend uses the software existing knowledge?	[Per94]
J-III.7	To what extend is the wording familiar?	[LCS97]
J-III.8	To what extend is graphic data connected with symbols?	[LCS97]
J-IV	Transparency	
J-IV.1	How is a way to preview "what will happen" provided?	[Per94]
J-IV.2	To what extend is the software not understandable?	[LHS08]
J-IV.3	To what extend is the effect of actions transparent?	[GH99]
J-IV.4	When did the software behave in an incomprehensible way?	[Kir96]
M	Memorability	
M.1	How is the memory load minimized by the software?	[Nie93]
M.2	How many details have you to remember?	[NM90, Per94]
M.3	How easy is it to find the important commands and actions?	[GH99, Prü97]
M.4	How easy is it to find the important commands and actions?	[GH99]
M.5	How easy is it to remember trained sequences?	[Prü97, Lun01, GH99]
M.6	How memorable is the naming of the menu items?	[LCS97, CDN88]
M.7	To what extend are data items kept short?	[LCS97]
M.8	To what extend verbal labels for icons are provided?	[LCS97]
M.8	When are prior answers recapitulated by the software?	[LCS97]
M-I	Consistency	
M-I.1	How is the integration of the various functions done?	[NM90]
M-I.2	How is the unambiguousness of terms and actions guaranteed?	[Bro96, GH99]
M-I.3	How is the consistency of control and action guaranteed?	[Per94]
M-I.4	How is the consistency of control and action guaranteed?	[LCS97, GH99]
M-I.5	How consistent is the interaction with the user?	[LCS97]
M-I.6	How consistent are the different views in the software?	[LCS97]
M-I.7	How consistent is the format of the display?	[LCS97]
M-I.8	How can related data entered with this software?	[LCS97]
M-I.9	How consistent is the formatting of the data fields?	[LCS97]
M-I.10	In what manner can the tasks performed with this software?	[Kir96, CDN88]
M-I.11	To what extend is the option wording consistent?	[LCS97]
M-I.12	To what extend is the software consistent to standards?	[Per94]
M-I.13	To what extend is the labeling of items consistent?	[LCS97, CDN88, GH99]
M-I.14	To what extend is the orientation of the display consistent?	[LCS97]
M-I.15	To what extend is the positioning of messages consistent?	[CDN88, GH99]
M-I.16	To what extend is the general design consistent?	[GH99, Prü97]
M-I.17	To what extend is the interaction paradigms consistent?	[Prü97]
M-I.17	What inconsistencies did you notice in this software?	[Lun01, Kir96, Bro96]
M-II	Customizability	

Id	Qualitative statement	Reference
M-II.1	How is the customization of windows facilitated?	[LCS97, GH99]
M-II.2	How is the individual assignment of command names done?	[LCS97]
M-II.3	How is the matching of software and user's skills done?	[LCS97]
M-II.4	How are recurring processes supported?	[Prü97]
M-II.5	To what extend meets the software your needs?	[Lun01]
M-II.6	To what extend is the software suitable for regular users?	[Lun01, LCS97, CDN88]
M-II.7	To what extend is the software suitable for occasional users?	[Lun01, LCS97, CDN88]
M-II.8	To what extend adapts the software to your knowledge?	[GH99, Prü97]
M-II.9	To what extend are the user's needs taken into consideration?	[Kir96]
M-II.10	To what extend is the data entry flexible?	[LCS97]
M-II.11	To what extend is the sequence control flexible?	[LCS97, Prü97, GH99]
M-II.12	To what extend can the speed of the software adapted?	[GH99]
M-II.13	To what extend can the software dynamically extended?	[Prü97]
M-II.14	To what extend can the software's output adapted?	[GH99, Prü97]
M-II.15	To what extend adopts the software the user's viewpoint?	[Prü97, GH99, Per94]
M-II.16	What do you think about the software's flexibility?	[Dav89, Lun01, CDN88]
M-II.17	When was the way you arrange your work disrupted?	[Kir96]
M-II.18	Which parts of the software are rigid?	[GH99]
M-III	Complexity	
M-III.1	How complex is the software?	[Bro96]
M-III.2	How complicated is the usage of the software?	[LHS08, Prü97]
M-III.3	How easy is it to forget how to do things with this software?	[Kir96]
M-III.4	How is the straightforward solution of problems supported?	[HBK03]
M-III.5	How dependent are you on assistance?	[Kir96]
M-III.6	To what extend do you limit yourself on a few commands?	[Kir96]
M-III.7	To what extend is the functionality limited?	[LCS97]
M-III.8	To what extend can you remember complex functions?	[GH99]
M-III.9	What information do you have to remember across dialogs?	[Per94]
M-III.10	Which parts of the software are simple?	[HBK03]
M-III.11	With which parts of the software are you unfamiliar?	[Kir96]
M-IV	Perspicuity	
M-IV.1	How is selected data highlighted?	[LCS97]
M-IV.2	How are long data items partitioned?	[LCS97]
M-IV.3	How are the available options communicated to the user?	[Kir96]
M-IV.4	How is the cursor made distinctive?	[LCS97]
M-IV.5	How are the display elements made distinctive?	[LCS97]
M-IV.6	How is direct comparison of items facilitated?	[LCS97]
M-IV.7	How do menus differ from other displayed information?	[LCS97]
M-IV.8	How is information grouped and organized?	[LCS97, CDN88]
M-IV.9	How is the readability of characters guaranteed?	[CDN88]
M-IV.10	To what extend is the information cluttered?	[LHS08]
M-IV.11	To what extend is the wording clear?	[LCS97]
M-IV.12	To what extend is the grouping and ordering of item logical?	[LCS97, Kir96]
M-IV.13	To what extend is the presented information understandable?	[Kir96]
M-IV.14	To what extend are the abbreviations distinctive?	[LCS97]
M-IV.15	To what extend have the commands distinctive meanings?	[LCS97]
M-IV.16	To what extend can the colors be distinguished?	[LCS97]
M-IV.17	To what extend are data fields visually distinctive?	[LCS97]
M-IV.18	To what extend is the information density reasonable?	[LCS97, Lew95, GH99]
M-IV.19	To what extend is the sequence of screens clear?	[CDN88]
M-IV.20	What parts of the software are confusing?	[LHS08, HBK03, CDN88]
L	Learnability	
L.1	How easy is it to learn to work with the software?	[Nie93]
L.2	How easy is it to become skillful with the software?	[Dav89, CDN88, GH99]
L.3	How many things had you to learn before using this software?	[Dav89, GH99, Lew95]
L.4	How many problems had you during the learning phase?	[Bro96]
		[Kir96]

Id	Qualitative statement	Reference
L.5	How long does it take to learn the software commands?	[Kir96, Lun01]
L.6	How difficult is the learning of new functionalities?	[Kir96, Lun01]
L.7	How is the learning supported by the software structure?	[LCS97]
L.8	To what extend is no-penalty learning supported?	[LCS97]
L.9	To what extend can users quickly learn to use this software?	[Bro96, Prü97]
L.10	To what extend had you to ask other persons for help?	[GH99]
L.12	What parts of the software are difficult to learn?	[LHS08]
L-I	Error handling	[Nie93]
L-I.1	How are errors prevented?	[NM90, Per94]
L-I.2	How hinders the software the user from making errors?	[Per94]
L-I.3	How is the error recovery performed?	[NM90, Per94, Lun01]
L-I.4	How can you leave an undesirable state?	[Per94]
L-I.5	How can operations canceled?	[LCS97]
L-I.6	How can errors corrected in this software?	[LCS97]
L-I.7	How is the repeated occurrence of an error indicated?	[LCS97]
L-I.8	How can the software restarted?	[LCS97, Kir96]
L-I.9	How can the user take back actions?	[LCS97, GH99]
L-I.10	How is the user warned against dangerous actions?	[GH99]
L-I.11	How is distinguished between feedback, warnings and errors?	[GH99]
L-I.12	How is guaranteed that no information gets lost?	[GH99]
L-I.13	How often lead errors to crashes of the software?	[GH99]
L-I.14	To what extend does the software assume that errors are made?	[Per94]
L-I.15	To what extend are error messages non-disruptive?	[LCS97, GH99]
L-I.16	To what extend can mistakes corrected?	[CDN88]
L-I.17	To what extend can the software be explored by trial and error?	[CDN88]
L-I.18	To what extend are the error messages helpful?	[CDN88, Prü97, Lew95]
L-I.19	To what extend have error minor consequences?	[Prü97, Lew95, GH99]
L-I.20	To what extend information on error recovery is provided?	[Prü97, GH99]
L-I.21	To what extend is the aborting of processes possible?	[GH99, Prü97]
L-I.22	What error messages are provided?	[NM90, Per94, LCS97]
L-I.23	What solutions are available in case of an error?	[Per94]
L-I.24	When is the user informed about wrong inputs?	[Prü97]
L-II	Feedback	[NM90, Per94]
L-II.1	How timely is the feedback about the processes?	[Per94, GH99]
L-II.2	How are erroneous entries displayed?	[LCS97]
L-II.3	How is feedback for control entries provide?	[LCS97]
L-II.4	How is the completion of processing indicated?	[LCS97]
L-II.5	How is the current position in menu structure indicated?	[LCS97]
L-II.6	How does the computer inform about its progress?	[CDN88]
L-II.7	How is the active window indicated?	[LCS97]
L-II.8	How is the feedback linked with the actions?	[GH99]
L-II.9	How is the visual feedback designed?	[GH99]
L-II.10	In which cases is no feedback available?	[Prü97]
L-II.11	To what extend is the feedback consistent?	[LCS97]
L-II.12	To what extend are instructions and prompts helpful?	[Kir96]
L-II.13	To what extend simplifies the highlighting the tasks?	[CDN88]
L-II.14	To what extend is the feedback understandable?	[GH99]
L-II.15	To what extend is the feedback adapted to the situation?	[Prü97]
L-III	Help	[NM90, Per94]
L-III.1	How are additional explanations accessed?	[GH99]
L-III.2	How task-oriented is the help provided by the software?	[Per94]
L-III.3	How has the software helped you to overcome your problems?	[Kir96]
L-III.4	How is help provided?	[LCS97]
L-III.5	How useful is the help information given by this software?	[Kir96]
L-III.6	How is an index of commands displayed?	[LCS97]
L-III.7	How is an index of data displayed?	[LCS97]

Id	Qualitative statement	Reference
L-III.8	To what extend varies the amount or quality of help?	[Kir96]
L-III.9	To what extend is the documentation informative.	[Kir96]
L-III.10	To what extend aids for entering data are provided?	[LCS97, GH99]
L-III.11	To what extend are the help messages useful?	[CDN88]
L-III.12	To what extend are the reference materials clear?	[CDN88]
L-IV	User Guidance	
L-IV.1	How does the software support the learning of the software?	[GH99]
L-IV.2	How can you explore new functions?	[GH99]
L-IV.3	How are all available possibilities displayed?	[GH99]
L-IV.4	How is an overview on valid inputs provided?	[GH99]
L-IV.5	How much do you have to read before using the software?	[Kir96]
L-IV.6	How usable is the software without documentation?	[Pri97]
L-IV.7	To what extend overview maps and trails are provided?	[Per94]
L-IV.8	To what extend is the wording of the user guidance consistent?	[LCS97]
L-IV.9	To what extend can you act on basis of the information?	[Kir96]
L-IV.10	To what extend is the user guidance flexible?	[LCS97]
L-IV.11	To what extend is the format of user guidance distinctive?	[LCS97]
L-IV.12	To what extend examples are provided beside explanations?	[GH99]
L-IV.13	To what extend situation specific explanations are provided?	[Pri97]
L-IV.14	When did you not know what to do next with this software?	[Kir96]
L-IV.15	When was the guidance information not available?	[LCS97]
L-IV.16	When had you to go back to look at the guides?	[Kir96]
P	Personal Effect	
P.1	To what extend is the software leading?	[LHS08]
P.2	What makes the software user friendly?	[LHS08]
P-I	Attractiveness	
P-I.1	To what extend is this software wonderful?	[Lun01]
P-I.2	To what extent is this software pleasant to use?	[Lun01, Lew95, LHS08]
P-I.3	To what extend is this software professional?	[HBK03]
P-I.4	To what extend is this software premium?	[HBK03]
P-I.5	To what extend is this software awkward?	[Kir96]
P-I.6	What is the value of this software?	[LHS08]
P-I.7	Which part of the software is stylish?	[HBK03]
P-I.8	Which part of this software is wonderful?	[CDN88, LHS08]
P-I.9	Which part of this software is stimulating?	[CDN88]
P-I.10	Which part of the presentation is attractive?	[Kir96, ?]
P-I.11	Why is this software interesting?	[LHS08]
P-I.12	Why is this software presentable?	[HBK03]
P-I.13	Why do you need to have this software?	[Lun01]
P-II	Novelty	
P-II.1	How is this software designed for advancement?	[Per94]
P-II.2	To what extend is this software creative?	[HBK03]
P-II.3	To what extend is this software dull	[LHS08]
P-II.4	What are the novel parts in this software?	[HBK03]
P-II.5	Which part of this software is innovative?	[LHS08, HBK03]
P-III	Satisfaction	
P-III.1	How comfortable do you feel using this software?	[Nie93]
P-III.2	How frustrating is the interaction with the software?	[Lew95]
P-III.3	How satisfied are you with the support information?	[CDN88]
P-III.4	How frequent would you like to use this software?	[Lew91]
P-III.5	How connects this software you with other people?	[Bro96]
P-III.6	To what extend is this software mentally stimulating?	[HBK03]
P-III.7	To what extend is working with this software satisfying?	[Kir96]
P-III.8	To what extend is this software exiting?	[Kir96, Lun01, Lew91]
P-III.9	To what extend are you satisfied with the ease of use?	[LHS08]
P-III.10	What makes the work with this software motivating?	[Lew95, Lew91]
		[LHS08]

Id	Qualitative statement	Reference
P-III.11	Which part of this software do you enjoy?	[Kir96, Lun01, Lew95]
P-III.12	Why would you recommend this software to your colleagues?	[Kir96, Lun01]
P-IV	Stress	
P-IV.1	How mentally demanding is the software?	[HS88]
P-IV.2	How physically demanding is the software?	[HS88]
P-IV.3	How temporally demanding is the software?	[HS88]
P-IV.4	How insecure, discouraged, irritated and stressed are you?	[HS88]
P-IV.5	How makes this software you a headache on occasion?	[Kir96]
P-IV.6	To what extend is the use of this software effortless?	[Lun01]
P-IV.7	To what extend is the use of this software frustrating?	[Kir96]
P-IV.8	When did you feel tense because of using this software?	[Kir96]
P-IV.9	When tends the software to be quiet?	[CDN88]

References

- [Bro96] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, Seiten 189–194, 1996.
- [CDN88] J.P. Chin, V.A. Diehl und K.L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 213–218. ACM New York, NY, USA, 1988.
- [Dav89] F.D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(3):319–340, 1989.
- [GH99] G. Gediga und K.C. Hamborg. IsoMetrics: An usability inventory supporting summative and formative evaluation of software systems. *HCI (1)*, Seiten 1018–1022, 1999.
- [GHD99] G. Gediga, K.C. Hamborg und I. Düntsch. The IsoMetrics usability inventory. *Behaviour & Information Technology*, 18(3):151–164, 1999.
- [HBK03] M. Hassenzahl, M. Burmester und F. Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer*, Seiten 187–196, 2003.
- [HS88] S.G. Hart und L.E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1:139–183, 1988.
- [KDSR07] U. Kuckartz, T. Dresing, C. Stefer und S. Rädiker. *Qualitative Evaluation: der Einstieg in die Praxis*. Springer, 2007.
- [Kin98] E.U. Kinast. Evaluation interkultureller Trainings. *Lengerich: Pabst*, 1998.
- [Kir96] J. Kirakowski. The software usability measurement inventory: background and usage. *Usability evaluation in industry*, Seiten 169–178, 1996.
- [LCS97] H.X. Lin, Y.Y. Choong und G. Salvendy. A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & Information Technology*, 16(4):267–277, 1997.

- [Lew91] J.R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. *ACM SIGCHI Bulletin*, 23(1):78–81, 1991.
- [Lew94] J.R. Lewis. Sample sizes for usability studies: Additional considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2):368–378, 1994.
- [Lew95] J.R. Lewis. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.
- [LHS08] B. Laugwitz, T. Held und M. Schrepp. Construction and Evaluation of a User Experience Questionnaire. *HCI and Usability for Education and Work*, Seiten 63–76, 2008.
- [Lun01] A. Lund. Measuring Usability with the USE Questionnaire. *Usability & User Experience*, 8(2), 2001.
- [May93] P. Mayring. Einf
ührung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken., 1993.
- [May08] P. Mayring. *Die Praxis der Qualitativen Inhaltsanalyse*. Beltz, 2008.
- [Nes08] S. Nestler. *Einsatzorganisation in Katastrophen. Unterstützung von Einsatzkräften durch mobile User-Interfaces*. VDM Verlag Dr. Müller, Januar 2008.
- [Nes10] S. Nestler. *Design, Implementation and Evaluation of User-Interfaces for life-threatening, time-critical and unstable Situations*. Dissertation, Technische Universität München, 2010.
- [Nie93] J. Nielsen. Usability engineering. *AP Professional, Boston*, 1993.
- [Nie94a] J. Nielsen. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, Seite 158. ACM, 1994.
- [Nie94b] J. Nielsen. Usability inspection methods. In *Conference companion on Human factors in computing systems*, Seiten 413–414. ACM, 1994.
- [NL93] J. Nielsen und T.K. Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, Seiten 206–213. ACM, 1993.
- [NM90] J. Nielsen und R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, Seiten 249–256. ACM, 1990.
- [Per94] G. Perlman. Practical usability evaluation. In *Conference companion on Human factors in computing systems*, Seiten 407–408. ACM, 1994.
- [Prü97] J. Prümper. Der Benutzungsfragebogen ISONORM 9241-10: Ergebnisse zur Reliabilität und Validität. In *Software-Ergonomie*, Jgg. 97, Seiten 253–262, 1997.
- [SD08] M. Sengpiel und D. Dittberner. The computer literacy scale (CLS) for older adults—development and validation. *Mensch & Computer 2008: Viel Mehr Interaktion*. 2008.
- [Vir92] R.A. Virzi. Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4):457–468, 1992.