

Usability Evaluation of Universal User Interfaces with the Computer-aided Evaluation Tool PROKUS

Gert Zülch and Sascha Stowasser

ifab-Institute of Human and Industrial Engineering, University of Karlsruhe,
Kaiserstrasse 12, D-76128 Karlsruhe, Germany
Email: ifab@mach.uni-karlsruhe.de

Abstract

The rapid dissemination of computer systems and accessible information systems (e.g. internet) is still unbroken. On the individual level, human-computer interfaces and dialogues change many people's lives: examples are the daily work at visual display terminals, digital lectures on the web and home working software tools. A few years ago, it was unusual to assess and evaluate such software or interfaces. Nowadays, researchers and practitioners are recognising the problems which grow up when usability evaluations are neglected. Usability evaluations can be carried out in different phases at the software lifecycle. It requires appropriate procedures with various techniques and methods to request and scope these different occasions.

For the purpose of systematic usability evaluation, the ifab-Institute of Human and Industrial Engineering of the University of Karlsruhe (Germany) has developed an evaluation tool called PROKUS (Programmsystem zur Kommunikationsergonomischen Untersuchung rechnerunterstützter Verfahren). This evaluation tool can be used for designing and carrying out usability evaluations in a reproducing and numerical process adapted to different occasions of the software lifecycle. The following paper will describe the evaluation process with PROKUS. Moreover, it will awake special attention to the scientific background of the development and the realisation of the evaluation tool PROKUS. A recently completed evaluation study on universal human-computer interfaces with PROKUS will complete the paper. For illustrating the procedure of PROKUS two web browsers exemplarily tested to what extent it fulfils the required standards of ISO 9241-14, 16 and 17.

1. NECESSITY AND OCCASIONS TO EVALUATE THE USABILITY OF HUMAN-COMPUTER INTERFACES

Due to the increasing use of data processing in all fields in human life (e.g. work, education, leisure) the attention given to a universal user-friendly software design is rising. The quality and acceptance of an interactive human-computer system is highly dependant upon the ergonomic design of the user interface. In the past, ergonomic aspects were often neglected when developing software and information systems. Thus, many users complain about the troubles in learning to use a software product or its insufficient functionality and complexity of interactions. Once the user becomes dissatisfied with a software product, a dialogue system or an accessible interface, this dissatisfaction dominates the functionality and the productivity of the system. A few years ago, it was unusual to assess and evaluate such software or interfaces. Nowadays, researchers and practitioners are recognising the problems which surface when usability evaluations are neglected. Furthermore, the evaluation of accessible interfaces (e.g. web pages) is becoming more and more important. New standards for multimedia interfaces, web pages or further advanced information and data technologies (virtual reality etc.) have to be developed and have to be transferred into the design of these modern human-computer systems.

The analysis and evaluation of software products or accessible interfaces, which should examine the adequate transfer of user-friendly standards to the design and development of products, can be realised at several occasions in the software lifecycle (figure 1). The first occasion is during the development process of the product by choosing an adequate human-computer interface. Quality tests assess mainly the functionality of the software. Conformity tests compare the product with the requirements of the standards and evaluate the user-friendliness of interfaces and dialogues. Another occasion, performed by the customer or the future user, is helping to select a suitable product by examining existing products on the market and comparing these products (comparison test). In this product selection process, conformity tests are made to assess the adherence to user-friendly standards. When the software has already been selected, several usability tests with additional goals and needs can be carried out in order to support an eventual redesign phase.

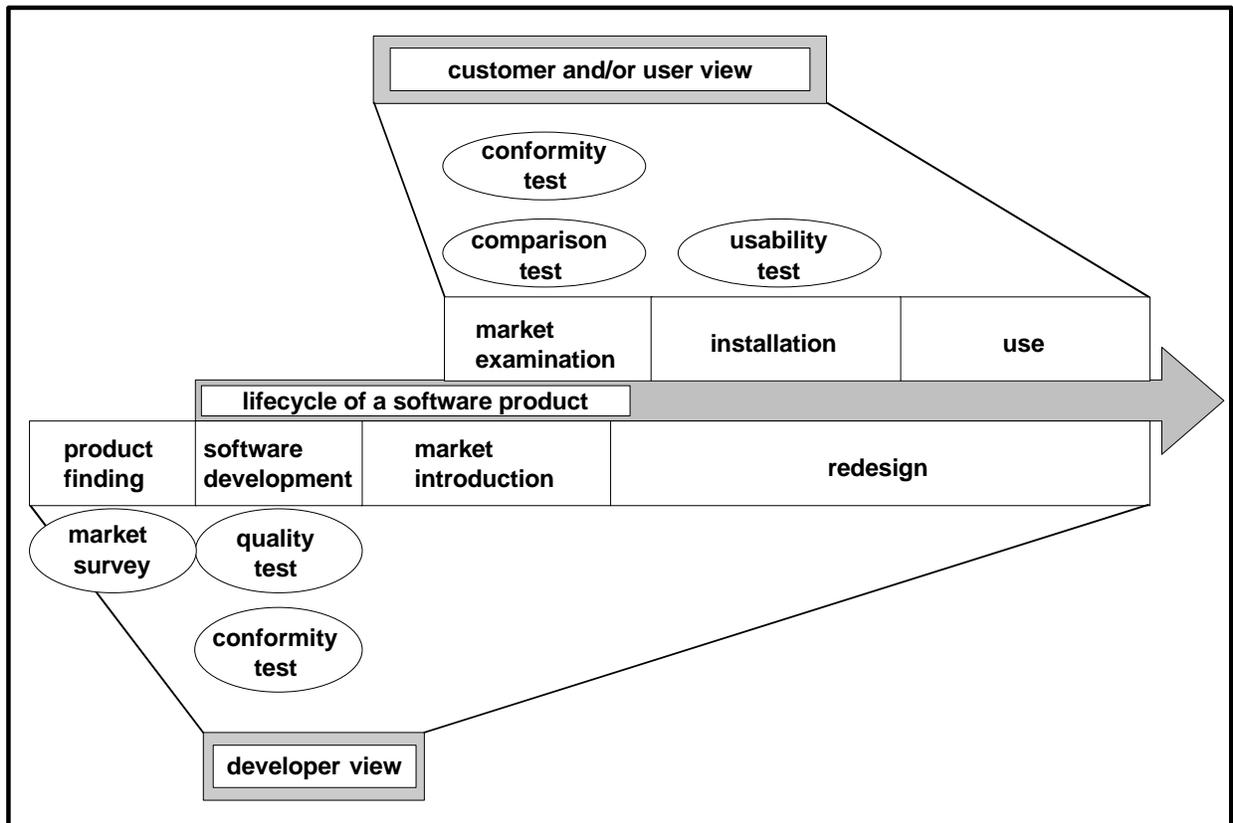


Figure 1. Occasions to evaluate software (following Zülch & Stowasser, 2000, p. 2/17; Zülch, Englisch & Grundel, 1993, p. 268)

2. THE COMPUTER-AIDED EVALUATION TOOL PROKUS

PROKUS (Programmsystem zur kommunikationsergonomischen Untersuchung rechnerunterstützter Verfahren), which has been developed at the ifab-Institute of Human and Industrial Engineering of the University of Karlsruhe, is a computer-system for the design of evaluation procedures and the performance of usability evaluations according to different evaluation situations. This tool can be classified as an evaluation method with guidelines and is applicable for systematic market examinations, conformity tests, quality tests and comparison tests of various kinds of software. This evaluation tool can be used to design and perform usability evaluations in a reproduction and numerical process adapted to different occasions of the software lifecycle.

2.1 System Elements of PROKUS

PROKUS is based on a catalogue with questions which are to be filled-in by an expert during the evaluation procedure. A central element of PROKUS is the exerciser database which consists of different series of investigations (figure 2). Each series of investigations contains several examinations which consist of a number of questions. Each question is described using the elements "criterion", "component", "task", "method", "importance" and "rating scale". The element "criterion" of the question represents the focused usability criterion. This criterion can be exemplarily derived from the seven general ergonomic principles (e.g. conformity with user expectations) which are described in ISO 9241-10.

The element "component" represents essential characteristics of the software or the interface which is to be evaluated with the respective question. One approach to arrange the components is the IFIP model for user interfaces developed by a working group of the German Informatics Community (GI) (IFIP stands for International Federation of Information Processing; Dzida, 1987, p. 339). As an example, the components of human-computer systems include user, task and computer components with input/output, dialogue and tool components. The element "task"

describes the function or purpose of the software or the interface according to the evaluated characteristic of the human-computer system (e.g. the button F7 is consequently used to save the data by the user). The element "method" represents the various test and evaluation methods with which the expert can evaluate the software and measure the required data in order to answer the question. ISO 9241-14 recommends the application of the evaluation methods "measurement", "observation", "expert judgement", "documental evidence" and "user testing". Regarding the special occasion of the evaluation, the use of a special method may be required. Evaluation with these methods means, that the evaluator tests an existing system and measures the required data or he derives these data from existing documents. The element "importance" presents the possibility of defining a ranking of the questions (e.g. class of question "1" means a "very important" question). The element "rating scale" represents the answering field for the actual question. Depending on the type of questions, PROKUS offers answering fields for three different scales, namely the nominal, ordinal and interval scales. An exact description of the evaluation system and the evaluation process with PROKUS can be read in Zülch and Stowasser (2000, p. 5/17).

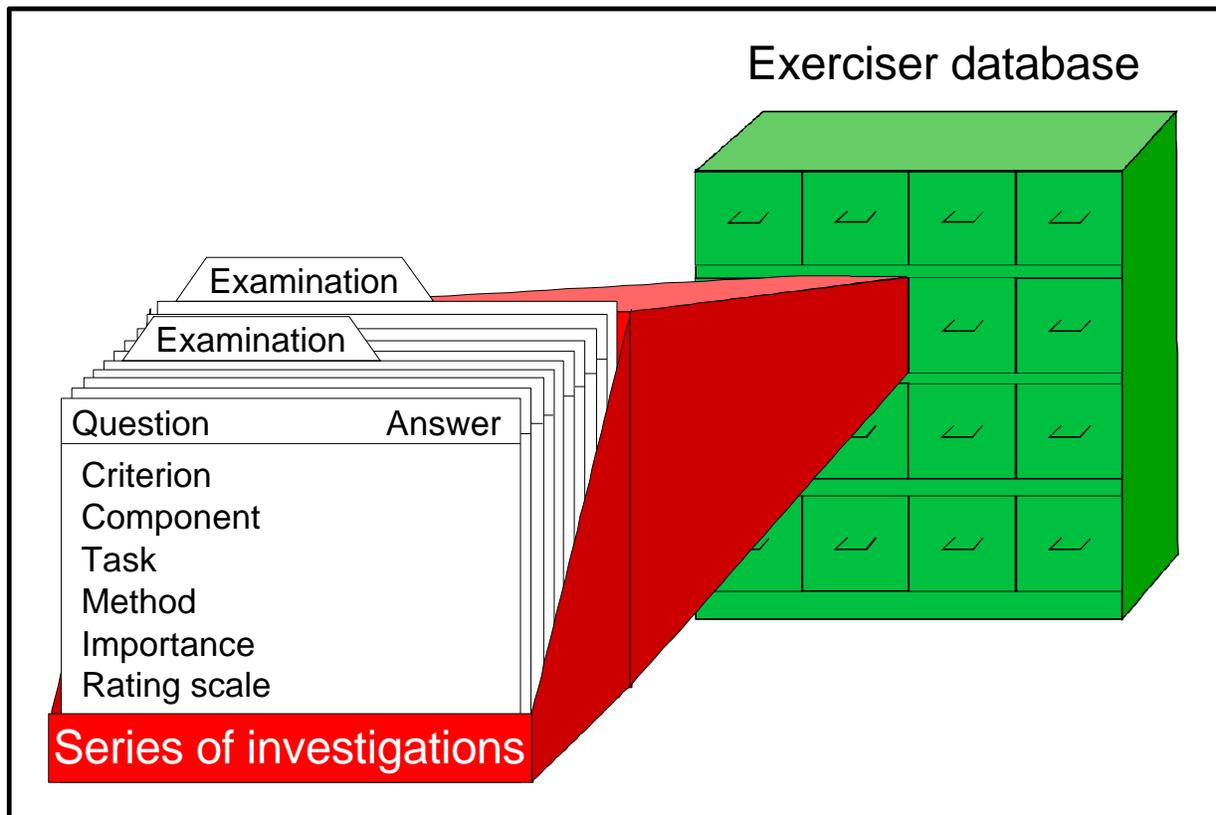


Figure 2: The evaluation tool PROKUS

2.2 Benefits in evaluating with PROKUS

Many procedures with different techniques and methods already exist for the evaluation of human-computer work systems and each evaluation refers to an adaptation to a special situation. For this purpose, PROKUS has been developed to design evaluation procedures according to different evaluation environments, e.g. resources or levels of detail, and to carry out different processes in usability evaluations. The complexity of an evaluation procedure designed with help from PROKUS can range from expert reviews and acceptance tests up to usability tests in laboratories.

A continual task in the conception of evaluation systems is the combination of the evaluation system with a system that stores different design guidelines, standards or other aspects concerning the evaluating matter. The evaluation tool PROKUS supports evaluators, designers and users on one hand with practicable and testable guidelines for human-computer interface design and on the other hand with methodological procedures for collecting evaluation and applying questions. These questions can be selected and combined in such a way which is desired by the

evaluator (of course a combination of questions of different standards, guidelines etc. is possible). For combining several questions PROKUS offers different sorting functions: e.g. choosing questions corresponding to just one evaluation method, questions for evaluating specific aspects (for instance only questions which can be applied to testing the dialogue component).

3. APPLICATION OF PROKUS FOR WEB BROWSER USABILITY STUDIES

The pursuit of internet connections has lead millions of users to navigate through the variety of web pages, join mailing lists, visit chat rooms, and fill newsgroups with useful information. Many results from other user-interface topics (e.g. menu selection, direct manipulation, screen design) can be applied to web-site design. However controlled experimental studies are necessary to evaluate the universal human-oriented internet performance (Shneiderman, 1998, p. 581). A recently completed evaluation study of universal human-computer interfaces web browsers carried out with PROKUS has demonstrated an application of this evaluation tool.

3.1 The evaluation process

Two common web browsers (called anonym as Browser A and Browser B) have been tested with an extensive series of usability evaluations. Both tests were performed independently: they were not carried out simultaneously. For each conformance test a series of investigations had to be defined (compare chapter 2.1). This series of investigations contains several examinations for a conformity testing in correspondence with ISO 9241 (Parts 14, 16, 17). These examinations were again arranged by a series of questions which are stored in the exerciser database of PROKUS. The evaluation was performed with a total of 288 questions (63 criteria, 24 scales, 21 methods, 13 tasks). The user group consisted of 5 persons (3 students, 2 non-academic persons).

3.2 General results of the evaluation

The PROKUS interpretation system includes various possibilities to aggregate and interpret the evaluation data. One possible result of the conformity test is expressed as the level of compliance (*LCO*) with chosen standards, which is represented by the number of fulfilled questions divided by the number of tested questions as a percentage. Figure 3 summarises selected *LCO*s considering both web browsers. For example, both systems provide nearly excellent navigation options (like "forward arrows", "history-lists"). Another criterion is the structure of the menus: e.g. grouping options in a universal user-compatible way.

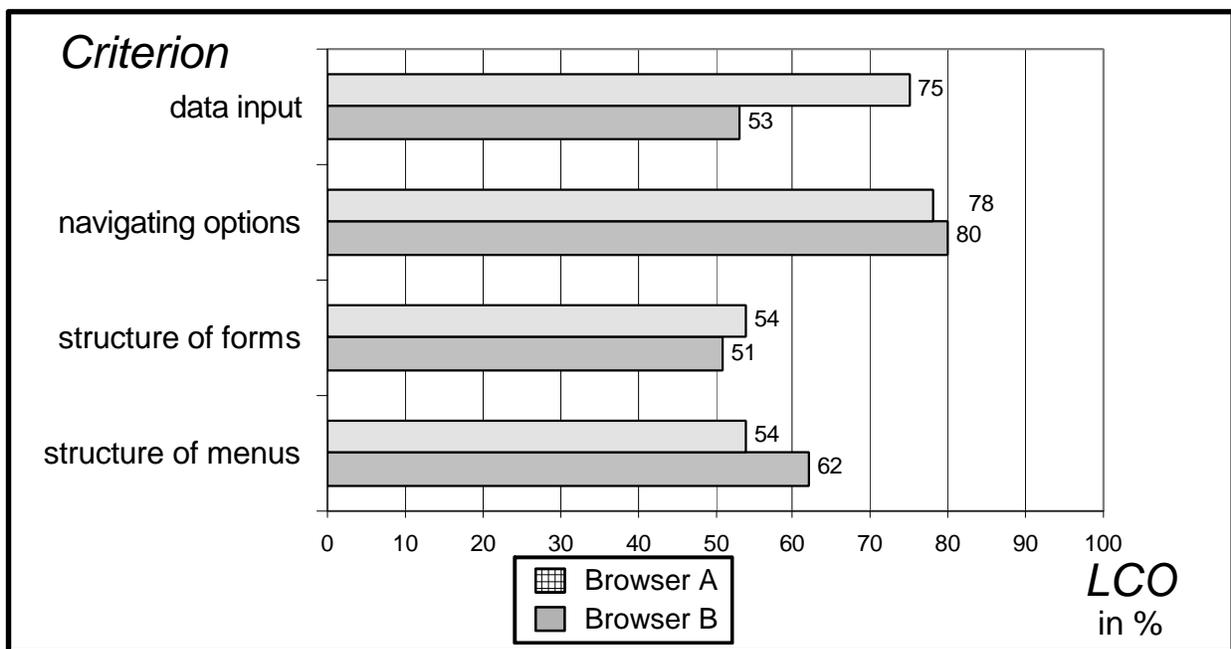


Figure 3: Level of compliance concerning selected criteria

The PROKUS interpretation system includes various possibilities to aggregate and interpret the evaluation data. One possible result of the conformity test is expressed as the level of compliance (*LCO*) with chosen standards, which is represented by the number of fulfilled questions divided by the number of tested questions as a percentage. Figure 3 summarises selected *LCOs* considering both web browsers. For example, both systems provide nearly excellent navigation options (like "forward arrows", "history-lists"). Another criterion is the structure of the menus: e.g. grouping options in a universal user-compatible way.

As a general result of the evaluation study, the overall level of compliance for Browser A is 69 %. Compared with this, Browser B measures 75 %. However, neither browser interface has a complete conformance with the considered standards. The results of the conformity test carried out with PROKUS shows that both interfaces need corrections in order to meet the required user-friendliness for universal application.

REFERENCES

Dzida, W. (1987). On Tools and Interfaces. In: M. Frese, E. Ulich, W. Dzida (Eds.), Psychological Issues of Human Computer Interfaces in the Work Place (pp. 339-355). Amsterdam: North-Holland.

ISO 9241-10 (1996). Ergonomics requirements for office work with display terminals (VDTs). Part 10: Dialogue Principles.

ISO 9241-14 (1997). Ergonomics requirements for office work with display terminals (VDTs). Part 14: Menu dialogues.

ISO 9241-16 (1999). Ergonomics requirements for office work with display terminals (VDTs). Part 16: Direct manipulation dialogues.

ISO 9241-17 (1998). Ergonomics requirements for office work with display terminals (VDTs). Part 17: Form filling dialogues.

Shneiderman, B. (1998). Designing the User Interface (3rd ed.). Reading, MA: Addison-Wesley.

Zülch, G., Englisch, J., & Grundel, C. (1993). Beurteilung der Benutzungsfreundlichkeit von Programmsystemen. Fortschrittliche Betriebsführung und Industrial Engineering, 42 (5).

Zülch, G., & Stowasser, S. (2000). Usability Evaluation of User Interfaces with the Computer-aided Evaluation Tool PROKUS. MMI-Interaktiv, 2 (3).