

Standards and software-ergonomic evaluation

Harald Reiterer^{a,b}, Reinhard Oppermann^a

^aGerman National Research Centre for Computer Science (GMD)
Institute for Applied Information Technology
P.O. Box 1316, D-53731 Sankt Augustin

^bUniversity of Vienna
Institute for Applied Computer Science and Information Systems
Liebiggasse 4/3-4, A-1010 Vienna

1. INTRODUCTION

Software-ergonomic evaluation is aimed at assessing a system's quality for particular users, for particular tasks, and in a particular environment. Functional, task oriented and user oriented issues are to be covered by an evaluation, where the user oriented perspective is the main focus of the consideration. An evaluation assesses a system against criteria. Criteria can be theory defined and measured in experiments where particular relationships between independent and dependent variables are constituted by the theory and tested by empirical data. Criteria can also be established by a political, industrial and scientific community for practical reasons, e.g., ending in standards. Standards are based on empirical evidence in HCI-research and practical and economical considerations. Not all desiderata from HCI-research are transferred into standards – e.g., there may be objections due to technical or financial constraints. The main advantage of standards are the contribution to an increased attention of the ergonomic subject in the community of developers and users of technology and the chance to refer to the described demands of the standards by the users of the technology. In some cases demand of standards are enforceable by law, e.g., to establish common working conditions for visual display terminals (VDT) users the European Union (EU) published a directive concerning the minimum safety and health requirements for VDT workers [EEC 1990]. The national governments of the EU members have to transform this directive into national law. The international standardisation activities in ISO 9241 concerning ergonomic requirements for VDTs are the basis to define the relevant requirements for the technology. An important consequence of these standardisation activities is that software developers and software buyers must take ergonomic requirements and principles into consideration. To assure the conformance of products with the established standards practicable software evaluation methods are needed.

2. STANDARDS

ISO 9241 "Ergonomic requirements for office work with visual display terminals (VDTs)" consists of the following parts to be considered during a software-ergonomic evaluation (the other parts are dedicated to task and hardware issues):

- Part 10: Dialogue principles
- Part 11: Guidance on specifying and measuring usability

- Part 12: Ergonomic requirements for presentation of information
- Part 13: User guidance
- Part 14: Menu dialogues
- Part 15: Command dialogues
- Part 16: Direct manipulation dialogues
- Part 17: Form-filling dialogues

ISO 9241 is far from being a pure technical standard that can be assured by quantitative measures. Rather it requires interpretation and tailoring to be useful in user interface evaluation and reflection of the state-of-the-art in research and development. It is subject to an ongoing process of discussion and negotiation. Different stages of consensus and ballots have to be achieved for the multipart standard. Different expertise and interest are influencing the results. They establish a "minimum level of user oriented quality" [Dzida 1995].

Although common terminology has been arranged for the multipart standard no general level of concreteness of the definition of the standard can be found. Diverse styles have been adopted in several parts: diverse structures and diverse levels of exhaustiveness. The concept of usability is defined in part 11 by effectiveness, efficiency and satisfaction of the user. The information presentation, i.e., the "look" of the interface, is described in part 12. The dialogue principles, i.e., the "feel" of the interface, is described in part 10. Requirements for user guidance, i.e. prompts, feedback, status, error support and help facilities, are described in part 13. The requirements for several dialogue techniques are described in part 14 to 17. Due to this structure there is a considerable amount of systematic overlap between parts of the standards, in particular between part 10 and 12 with part 14 to 17.

Part 10 establishes a systematic framework of ergonomic principles for the dialogue techniques with high-level definitions but with only illustrative applications and examples for the principles. All principles appear in the parts 13 - 17 of the standard but without explicit reference. Part 13 to part 17 of the standard define more or less low-level and fairly exhaustive requirements for the user interface. The structure follows a technological systematic of interface techniques and systems components. In many cases task-, user-, and technical environment aspects are considered as conditions of the applicability or relative relevance of the specified requirements. These aspects constitute the context of use defined in part 11 to be considered in applying the standard for a given work system (see the next section).

Part 11 gives the following definition of usability: "Usability is measured by the extent to which the intended goals of use of the overall system are achieved (effectiveness); the resources that have to be expended to achieve the intended goals (efficiency); and the extent to which the user finds the overall system acceptable (satisfaction)." For a more detailed discussion of the term usability see [Bevan 1995]. Effectiveness, efficiency and satisfaction can be seen as quality factors of usability. To evaluate these factors, they need to be decomposed into sub-factors, and finally, into usability measures. [Dzida 1995] presents a usability quality model that refines the term usability. This model introduces a stepwise operationalisation of the factors effectiveness, efficiency and satisfaction, which ends up with specific measures called criteria¹. Each criterion is defined as a required level of measure of a usability factor, the achievement of which can be verified. Example: The usability factor learnability can be measured with the help of the criterion: "the use of an application program has to be learned

¹ Note that in the author's notation the term criterion is used different from our terminology: we perceive a criterion as an abstraction of different measures for an ergonomic quality, e.g. flexibility, what Dzida calls sub-factor, where he uses a criterion for the measurable operationalisation of a sub-factor.

within 30 minutes". The criteria define measurable requirements and therefore the evaluation of usability implies a comparison between software product attributes and the measurable requirements. Another model to refine usability factors is the linguistic decomposition approach [Vanderdonck 1995].

3. CONTEXT OF USE

The software-ergonomic evaluation of usability has to be placed in the context of use consisting of the users, their jobs and tasks, the hardware and software, and the organisational, technical, and physical environment. Although usability is a property of the overall system, the focus of attention is usually on a specific element within the overall system – in our case the software product. It is possible to address the usability of the user interface, but only if the particular context of use has been identified. The investigation of the elements of the context of use is done by considering the following characteristics [ISO 9241 Part 11]:

- *The user*: User types (e.g., user populations) based on aspects about users skills and knowledge (e.g., software experience, hardware experience, task experience, organisational experience, education, training), personal attributes (e.g., age, gender, physical capabilities, disabilities), cognitive attributes (e.g. intellectual abilities, motivation).
- *The software*: Descriptions of the functionality and main application areas of the software, available instructional items (e.g. handbooks).
- *The job and tasks*: Details about the job of the user as a whole, and the tasks for which the software will be used as an aid (e.g., task goal, task name, task frequency, task breakdown, task duration, task flexibility, task output, task dependencies).
- *Organisational environment*: Aspects of the structure of the organisation (e.g. hours of work, group working, job function, work practices, management structure, communication structure, interruptions) the attitudes and culture (e.g., policy on use of computer, organisational aims, industrial relations), and the job design (e.g., job flexibility, performance monitoring, performance feedback, pacing, autonomy, discretion).
- *Technical environment*: Hardware and basic software (e.g. operating system) which is necessary to use the software, reference material.
- *Physical environment*: Workplace conditions (e.g., humidity, temperature, noise), design of the workplace (e.g., space and furniture, user posture, location), workplace safety (e.g. health hazards, protective clothing and equipment).

4. EVADIS - EVALUATION BASED ON STANDARD REQUIREMENTS

The GMD is developing in close co-operation with scientific (University of Vienna) and industrial partners (TÜV Rheinland, Cologne; Prümper & Partner, Munich; Ernst & Young, Vienna) an evaluation approach called Evaluation of Dialogue Systems (EVADIS III). The EVADIS III approach is based on EVADIS II [Oppermann et al. 1992]. The aim of the development of this version of EVADIS is to consider the requirements of the ISO 9241 standard in much more detail. The application of ISO 9241, part 8 and 10 to 17 is supported by the EVADIS procedure. To facilitate the application of the multipart standard explanations for the understanding of the requirements and explanations for the completion of conformance testing are provided. To provide these facilities, i.e., a systematic, explanations, and test instructions, are the aim of EVADIS III. Other examples of tools for evaluating the usability of user interfaces are described in [Balbo 1995].

4.1 Expert based Evaluation Approach

The EVADIS evaluation approach is a comprehensive one consisting of a combination of methods forming a method based expert judgement:

- a guideline-oriented checklist of test items to evaluate the user interface of the software,
- a questionnaire to explore user characteristics, and
- a simplified work place analysis to explore the context of use.

The evaluator is a human factors expert using the methods to evaluate the conformance with ISO 9241. The expert approach is based less on a task to be performed by the tested system than on questions asked by software ergonomics. This method is subjective since the expert examines and answers questions according to his personal assessment. It is objective since the ergonomic requirements are operationalized and precisely formulated to an extent enabling the evaluator to answer questions based on clear test rules and traceable conditions. Advantages of an expert based evaluation method are: relatively fast, uses few resources, provides an integrated view, and can be addressed to a wide range of behaviour.

4.2 Evaluation Procedure

Detailed test instructions in the EVADIS evaluation guide help to reduce the subjectivity of this method. The evaluation guide consists of five steps and includes a detailed evaluation process description with a clear notation and structure. The first three steps can be executed simultaneously. The results of these three steps are a list of attributes describing the user, the task, and the system context. Step 4 is the central step of the evaluation process. Here all selected test items have to be answered considering the context of use. The result of these activities is a test record that forms the basis for the interpretation of the results and the writing of the test report in step 5. The evaluation procedure can be used for summative or formative evaluation purposes. Therefore it has to be embedded in the software engineering process at different stages [Reiterer and Oppermann 1993]. [Hix 1995] discusses in more detail the incorporation of usability evaluation techniques into the software lifecycle.

4.3 Components of the Evaluation Procedure

The *simplified work place analysis* and the *questionnaire to explore user characteristics* are still under development. Their development will be based on the necessary context of use information (e.g., tasks attributes, user attributes, system attributes) which are derived during the development of the test items. This retrograde development process guarantees that only necessary information will be gathered with these methods and helps to develop a highly practicable evaluation procedure.

The guideline-oriented checklist of test items is at present based on the requirements of ISO 9241 part 8, 10, and 14 that have reached the Draft International Standard (DIS) status. Parts 12, 13, 15, 16, 17 will be included when they have reached the DIS status. The current work is concentrated on the development of the test items. They are embedded in a two-dimensional framework. The first dimension forms the technical system components – inspired by the IFIP model for user interfaces. The second dimension forms the software-ergonomic principles – inspired by the dialogue principles of ISO 9241 part 10. For a discussion of how usable usability principles are see [Bastien and Scapin 1995].

Figure 1 shows a sample item in more detail: how each item is typically structured and how the context of use is considered. To test a software product for conformance the standard requirements have to be interpreted regarding the criteria of usability that are derived from the

product's context of use [Dzida 1995]. In the example item the "if-clause" refers to the condition of application, the remaining sentence provides the guideline, i.e., recommendation according to the involved "should". Therefore the conformance test is a two-stage process. First the application of the "if-clause" is to be proven. If the condition is not applicable this will be marked under the test situation; otherwise a comparison of the product attributes with the criteria of usability will be done. In this comparison the context of use has to be considered. The necessary context information has been gathered using the simplified work place analysis and the questionnaire to explore user characteristics. The test instruction gives the evaluator useful information how to test the specific attribute of the software product systematically. The comment contains desirable ergonomic requirements. The evaluator can use this information during the rating process. Normally a test item will be answered multiple times in different test situations $i=1, \dots, n$. Each test situation will be shortly characterised by the evaluator (e.g., menu structure of the main menu; menu structure of the customer window). For each test situation a new rating will be done. Answering a test item in different test situations is necessary for a systematic evaluation of a software product. Different answers in different situations can also be an indication for an inconsistent user interface. Using the ratings of each item a final judgement can be made whether or not the software fulfils the ergonomic requirements of ISO 9241.

Component: 2.1.5.1 Menu structure	Criterion: Conformity with user expectation
Source: ISO 9241 Part 14	Item No.: 2151.04.10
Standard requirement:	
If options can be arranged into conventional or natural groups known to users, options should be organised into levels and menus consistent with that order.	
Context of use:	
Task context: Examine task specific conventions	
User context: Examine user specific conventions	
System context: -	
Test instruction:	
Examine all menu items concerning a consistent order of the structure.	
Comment:	
Menu structures should reflect user expectations and facilitate the user's ability to find and to select menu options relevant for the task and should support the user's flow of work. Task specific or user specific conventions are better than logical or arbitrary ones.	
Test situation $i=1, \dots, n$: _____	
Criteria of usability:	
<input type="radio"/> consistent order <input type="radio"/> no consistent order <input type="radio"/> not applicable	
Explanation of the rating:	

Figure 1. Example of a test item

4.4 Evaluation Software

An evaluation software available under Microsoft Windows supports the evaluator during the whole evaluation process and provides an assessment summary. The software presents all test items on the screen. The evaluator selects the criteria of usability and writes an explanation of the rating. The software calculates an average mark for each ergonomic criterion and each technical component and sorts the results by components or ergonomic principles. Therefore the evaluator is freed of routine work and can concentrate on the evaluation.

5. SUMMARY

There is an increasing need for practical and comprehensive evaluation methods and tools for conformance testing with standards. Practical means that the amount of time and resources must be manageable in software projects. Comprehensive means that the context of use has to be considered during the evaluation of user interfaces. The evaluation approach EVADIS III is such a practical and comprehensive one. In particular, it takes the context of use into consideration and provides computer support for the use of the evaluation procedure. It supports the evaluator during the evaluation process with detailed instructions. The first version of EVADIS III should be available at the end of 1995.

REFERENCES

- Balbo, S. 1995, Software Tools for Evaluating the Usability of Human-Computer Interfaces. In this volume.
- Bastien, Ch., Scapin, D. 1995, How usable are usability principles, criteria and standards? In this volume
- Bevan, N. 1995, What is usability? In this volume
- EEC, 1990, The minimum safety and health requirements for work with display screen equipment, Directive of the European Economic Community, 90/270/EEC
- Hix, D. 1995, Usability Evaluation: How Does It Relate to Software Engineering? In this volume
- Dzida, W. 1995, Standards for user-interfaces, in: *Computer Standards & Interfaces 17* (1995), 89-97
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 8, Requirements for displayed colours, Draft International Standard, June 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 10, Dialogue Principles, International Standard, September 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 11, Guidance on specifying and measuring usability, Committee Draft, May 1993.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 12, Ergonomic requirements for presentation of information, Committee Draft, October 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 13, User guidance, Draft International Standard, May 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 14, Menu dialogues, Draft International Standard, May 1994,
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 15, Command dialogues, Proposed Draft International Standard, April 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 16, Direct manipulation dialogues, Committee Draft, October 1994.
- ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals, Part 17, Form filling dialogues, Committee Draft, December 1994.
- Oppermann, R., Murchner, B., Reiterer, H., Koch, M. 1992, *Software-ergonomische Evaluation - Der Leitfaden EVADIS II*. Walter de Gruyter, Berlin, 1992.
- Reiterer, H., Oppermann, R. 1993, Evaluation of User Interfaces, EVADIS II - A comprehensive Evaluation Approach. *Behaviour & Information Technology 12* (1993), 3, 137-148.
- Vanderdonck, J. 1995, Using Recommendations and Data Base Tools for Evaluation by Linguistic Ergonomic Criteria. In this volume.