Software-Evaluation based upon ISO 9241 Part 10*

Jochen Prümper

Abstract. In this article the software evaluation instrument ISONORM 9241/10 is presented. This instrument was developed from 'ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs), Part 10: Dialogue Principles'. The capabilities of this instrument are shown through two methods. First, two different kinds of computer-systems (with and without graphical user interface) are evaluated for their conformity to ISO 9241/10. Second, a procedure is shown in which ISONORM 9241/10 is used as a basis for moderation with user-groups, when in the beginning of a design-process, through the forum of a participative system-development, first demands are formulated. It is shown that the user-friendliness of systems with GUI is judged to be significantly superior in all seven principles of ISO 9241/10, to those systems without GUI. In the user-group meeting, through ISONORM 9241/10, many concrete recommendations for a new software to be developed could be generated.

1 Introduction

After many years of preparation (e.g. DIN 66 234/8, 1988 [3]; Dzida & Itzfeldt, 1978 [4]) in february 1993 the 'International Organisation for Standardisation' (ISO) presented the first 'Draft International Standard' of 'ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs), Part 10: Dialogue Principles' (ISO 9241/10, 1993 [10]).

ISO 9241/10 deals with software aspects and describes seven general ergonomic principles, which are independent of any specific dialogue technique; i.e. they are presented without reference to situations of use, applications, environments or technology.

The following table gives an overview of the seven principles of ISO 9241/10 and its corresponding descriptions.

---

* This article was produced as part of the software-development project "Verlag 2000" ("Publication 2000"). "Verlag 2000" is supported by a grant from the Work & Technical Fund of the Ministry of Research and Technology of the Federal Republic of Germany (No.: 01 HK 601 8). Thanks are due to Michael Antt and Mark Farah.

Tab. 1. The seven principles of ISO 9241/10 and corresponding descriptions

Principle	Description
• suitability for the task	A dialogue is suitable for a task to the extent that it supports the user in the effective and efficient completion of the task.
• self-descriptiveness	A dialogue is self-descriptive to the extent that each dialogue step is immediately comprehensible through feedback from the system, or is explained to the user when requesting the relevant information.
• controllability	A dialogue is controllable to the extent that the user is able to maintain direction over the whole course of the interaction until the point at which the goal has been met.
• conformity with user expectations	A dialogue conforms with user expectations to the extent that it corresponds to the user's task knowledge, education, experience, and to commonly accepted conventions.
• error tolerance	A dialogue is error tolerant to the extent, if despite evident errors in input, the intended result may be achieved with either no or minimal corrective action having to be taken.
• suitability for individualization	A dialogue is suitable for individualization to the extent that the dialogue system is constructed to allow for modification to the user's individual needs and skills for a given task.
• suitability for learning	A dialogue is suitable for learning to the extent that it provides support and guidance to the user during the learning phases.

Although the type of information in ISO 9241/10 is more a general guidance (and thus has a more informative than normative character), the principles are intended to be used in the design as well as in the evaluation of dialogue systems (see Prümper, 1993b [16]; see also Nielsen, 1992 [13]; Nielsen & Molich, 1990 [14]).

2 The Software-Evaluation Instrument ISONORM 9241/10

In order to analyse whether a software-system meets the seven principles of ISO 9241/10, those principles must be characterized through an evaluation instrument.

In this article an evaluation instrument based on ISO 9241/10 will be presented. Going forward, the name of this instrument will be ISONORM 9241/10 (see Prümper & Anft, 1993b [19]; 1993c [20]).

The intention in developing ISONORM 9241/10 was to create a competent, practical and compact instrument which can be efficiently used to evaluate software and to support software development.

For this purpose the questionnaire method is the most pragmatic because of its inherent advantages of little effort and easy use. In order to meet the claim of an economically usable instrument, the extent of the questionnaire was limited to five items for each of the seven principles of ISO 9241/10.

For the construction of the questionnaire, we decided upon a seven tier, bi-polar question format with an answer schema from "- - -" to "+++" (coded: 1 - 7). Through the opposite positioning of the positive and negative poles we believe that the subjects are motivated to reflect upon the statements in more detail and are not influenced to answer towards a particular side.

Figure 1 shows an example item from the questionnaire regarding the principle "conformity with user expectations".

The software ...

	---	--	-	-/+	+	++	+++	
makes more difficult the orientation because of a non-conforming design.								simplifies the orientation because of a conforming design.

Fig. 1. Example item from the questionnaire regarding the principle "conformity with user expectations"

According to action theory (Frese & Sabini, 1985 [7]; Frese & Zapf, 1994 [9]; Hacker, Volpert & von Cranach, 1982 [10]), we tried to formulate the items as close as possible to operation and activity.

The language itself was chosen so that the specific properties of different interfaces are sufficiently differentiated, but at the same time is appropiate for many software applications. With this "medium level of granulation" (see Rauterberg, 1992a [22]) ISONORM 9241/10 offers the evaluation of a wide range of software systems. Table 2 shows an example item for each of the seven principles.

Tab. 2. The seven principles of ISO 9241/10 and respective item examples

Principle	Item
• suitability for the task	The software inappropriately meets the demands of the tasks.
• self-descriptiveness	The software offers insufficient information regarding the inputs which are allowed or necessary.
• controllability	The software forces an unnecessary inflexible sequence of commands.
• conformity with user expectations	The software makes more difficult the orientation because of a non-conforming design.
• error tolerance	The software gives unspecific information regarding error correction and management.
• suitability for individualization	The software is difficult for the user to expand if new tasks arise.
• suitability for learning	The software is difficult to learn without outside direction or handbooks.

However, an evaluation instrument which judges user-friendliness of software is, among other things, only useful if it is able to differentiate significantly between different software programs (this rises the question of the validation of an instrument). Furthermore, from its use practible implications should arise.

This is proved by the following: First, two different kinds of software interfaces are tested for norm-conformity through the use of ISONORM 9241/10. Second, a procedure will be demonstrated in which ISONORM 9241/10 was used during the beginning of a design-process as a basis for the moderation of user-groups.

3 Comparison of Software With and Without Graphical User Interface Using ISONORM 9241/10

A graphical user interface (GUI) can be defined as an interface that allows users to choose commands and other options by pointing to a graphical icon and then activating the choice by either the keyboard or a mouse.

In the literature there is evidence that systems with graphical user interface which are characterized by direct manipulation of objects are superior to conventional systems.

For example, following a description of the most important characteristics of direct manipulation (model action world, direct information and direct action) Altmann (1987) [1] and Frese, Schulte-Göcking and Altmann (1987) [8] showed that

learning progress was higher with a direct manipulation system than with a conventional system, and Téeni (1990) [26] demonstrated that direct manipulation enhanced *cognitive control* in a judgement task (the feedback generated by direct manipulation was more effective compared to distinct feedback in traditional human-computer dialogue). Ulich, Rauterberg, Moll, Greutmann & Strohm (1991) [27] conducted a comparison of a desktop interface (high transparency) and a conventional menu selection interface (low transparency) to prove the criterion of *transparency*, and demonstrated the superiority of the user interface with direct manipulation over the conventional user interface with menu selection, and Rauterberg (1992b) [23] showed in an comparison of menu-selection and desktop computer programs the superiority of the desktop user interface for beginners as well as for experts. Furthermore, Shneiderman (1988) [25] encouraged greater attention to direct manipulation in which the objects and actions are visible, the actions are invoked by selection or pointing, and the impact is immediately visible and reversible. Rosson & Alpert (1990) [24] - considering the potential of the object-oriented paradigm in easing the design process of interactive software design and improving the usability of the resulting system - came to the conclusion that object orientation allows a designer to generate an initial design model in the context of the problem itself, rather than requiring a mapping onto traditional computer science constructs.

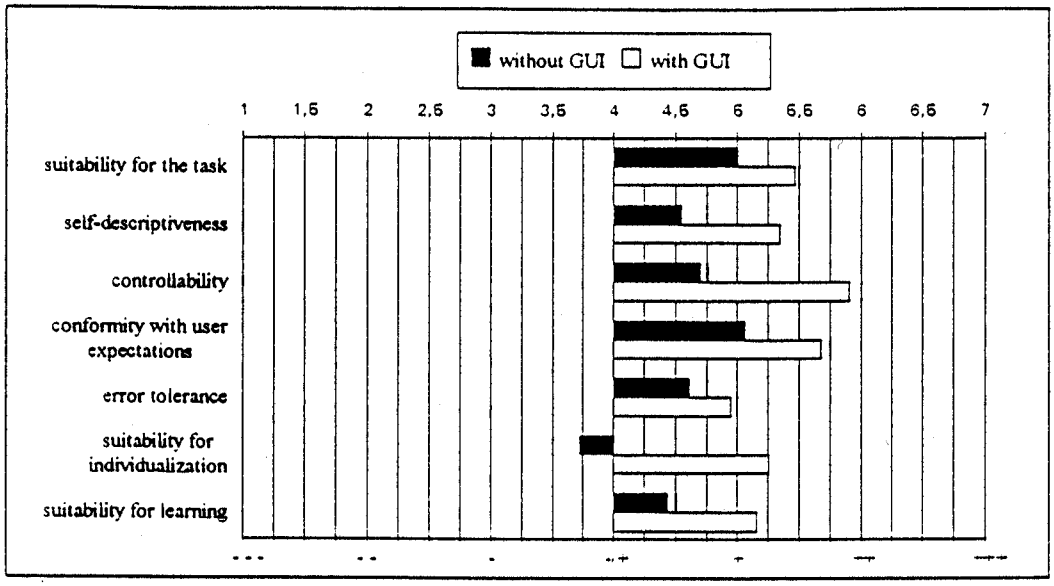
Thus, because of theoretical reasons and empirical results it can be hypothized that software with graphical user interface is more user-friendly, i.e. meets better the principles of ISO 9241/10, than software with no graphical user interface.

3.1 Procedure

Until now 350 users from 20 different companies in the Federal Republic of Germany participated in the evaluation of software using ISONORM 9241/10. They were selected through different methods, such as articles in magazines and journals, contacts to companies or participation in software-development projects. The average age of the subjects was 35 years old, 64,2 % were female, 35,8% male. Those 350 users evaluated 66 different software programs (without differentiating versions). 106 users evaluated 22 different software-programs with a graphical user interface (mainly diverse Microsoft-Windows and Apple-Macintosh Programs) and 244 users evaluated 44 different software-programs with no graphical user interface (mainly software running MS-DOS, IBM AS/400, IBM /36 or Siemens BS 2000).

3.2 Results

The following graphic shows a comparison of programs with graphical user interface to those without graphical user interface using the seven principles of ISO 9241/10 (see Figure 2).



Note: All differences between systems with graphical user interface and without graphical user interface are significantly at p< .001 (t-test, one-tailed probability); N=350

Fig. 2. Comparison of programs with graphical user interface to those without graphical user interface using the seven principles of ISO 9241/10

As can be seen from Figure 2, the user-friendliness of the programs with graphical user interface is significantly superior in all seven principles of ISONORM 9241/10 to the user friendliness of programs with no graphical user interface. This can be interpreted as first proof of a successful realization of 'ISO 9241: Ergonomic requirements for office work with visual display terminals, Part 10: Dialogue Principles'.

Moreover, the above results demonstrate that the user-friendliness of software in general is not yet to be judged as satisfactory. With a total-average for all programs of x=4,95 the user-friendliness of both the software with graphical user interface (x=5,38), and definitly the software without graphical user interface (x=4,52), are not very impressive. Only regarding the two principles "suitability for the task" (x=5,00) and "conformity with user expectations" (x=5,06) did the software without graphical user interface reach the "+" - mark; "suitability for individualization" (x=3,72) is even in the negative zone.

But even the user-friendliness of the software with graphical user interface still leaves much to be desired. In addition, none of these programs reached the "++" - mark at one principle ("controllability" with x=5,91 is rated the highest) and "error tolerance" didn't even reach the "+" - mark (x=4,61).

In conclusion, one can say that ISONORM 9241/10 is able to significantly distinguish between different types of software, and that the instrument can even point out areas for improvements of modern GUI software.

4 ISONORM 9241/10 in User-Oriented Software-Development

In the following, a procedure will be shown in which ISONORM 9241/10 was used as a basis for the moderation of user-groups, when in the beginning of a design-process first demands needed to be generated. This procedure took place in a software-development project for publishing-houses (see Prümper, 1993a [15]; 1993c [17]). In this project the software-development is managed in close participation with potential end-users. At the forefront, active participation is practiced where the users have the possibility to influence the design-process at its early stages. For this purpose end-users and software-developers come together from time to time under the guidance of industrial psychologists, in order to define the individual design-phases. At these meetings the later users are introduced to developing versions of the publishing software.

However, in most cases only a selected circle of end-users have the opportunity to take part in these meetings. Therefore, the rest of the future users will at the least get the status of passive participation. Although these users have no direct influence on the development of the software, their demands can also be taken into account. Here, ISONORM 9241/10 will be of usefull service.

4.1 Procedure

In order to realize active as well as passive participation, 42 users from three German publishing-houses evaluated their old "address"-module with ISONORM 9241/10 (passive participation). The average age of the subjects was 35 years old, 92,9 % were female, 7,1% male. From these publishers seven users participated in the first meeting of the user-group (active participation) to determine the existing weak points of the old "address"-module in order to generate suggestions for the new publishing-software.

At the beginning of the meeting item by item results of the inquiry were presented to the user-group and for every item examples were requested (with use of a flip-chart). Those examples were worked on in the form of group discussion and then fine-tuned. Hereby, the users had access to the software they had evaluated, to further demonstrate their examples. This procedure was repeated seven times according to the seven principles of ISO 9241/10. The time spent was seven hours (for details see Prümper & Anft, 1993a [18]).

4.2 Results

It wasn't always easy for the users to spontaneously generate an example for each item. In these cases, the moderator explained the items according to the "critical incident technique" (Flanagan, 1954 [6]) with positive and negative examples of other programs or asked the users to explore the system with smaller tasks. However, the fact that the users could not always precisely differentiate between the individual principles of ISO 9241/10 proved to be a difficulty. For example, some conceptual cases for "controllability" or "conformity with user expectations" should have been taken into account by the principle "suitability for the task". The reason for this

overlapping does not only lie on the user side, but rather are inherent in the principles of ISO 9241/10 themselves: "The dialog principles are not independent, and it may be neceessary to trade off the benifits of one priniple against another" (ISO 9241/10, 1993, iv) [11]. The consequences were that the users were able to create more example from the principles which were presented at the beginning of the moderation than those in the end. However, the final conclusion of this procedure was that the user-group supplemented the results of the ISONORM 9241/10 inquiry with a number of concrete suggestions. As an illustration of the procedure, the principles and respective results are presented in Table 3.

Tab. 3. The seven principles of ISO 9241/10 and exemplary results from the user-group

Principle	Fault of the Principle
• suitability for the task	The address-screen contains unnecessary fields, which are rarely if ever used.
• self-descriptiveness	The address-screen does not distinguish between fields which are definitely necessary or just optional.
• controllability	When defining a new address the user is forced to enter an unnecessary screen, even though under the circumstances it is not required.
• conformity with user expectations	In the address-module the function-keys perform differently as in the book-module.
• error tolerance	If an address is incompletely entered, the cursor moves to the first field in the screen and the user receives the error-message: "Entry false or incomplete". It is not obvious which field is intended.
• suitability for individualization	If new codes are necessary (i.e. hobbies or subject-groups), the user can not supplement the codes by him- or herself.
• suitability for learning	Learning the computer is strenuous because no on-line help exists.

Concerning "suitability for the task" one could say that the address-module does not offer all functions to fulfill the demands, and that unnecessary entries are required. Regarding "self-descriptiveness", it was criticised that no context-specific explana-

tions are available. Negative comments were also noted concerning "controllability". It was especially mentioned that the software doesn't offer the possibility to interrupt the work without loss. Moreover, the user can hardly influence which type of information is offered to him. Concerning "conformity with user expectations", it was criticised that the software is not uniformly designed so that it is difficult to transfer knowledge from one module to another. In the context of "error tolerance", it was noted that the software did not give concrete information for error correction and management, and that it didn't always give direct feedback regarding false entries. In the rating of "suitability for individualization" it was noted that the address-module is hardly expandable if new tasks arise, and that it is difficult to mould the software to the individuals' working preferences. Finally, the software is critically judged concerning "suitability for learning" because it doesn't easily adapt when trying out new functions. Moreover, there is no on-line help available, so the software is very difficult to learn without outside aid.

Clearly by all principles, the users of the address-module in consideration asked for many improvements.

5 Discussion

In this article an instrument for software evaluation based on the ISO 9241/10 was introduced, and its capabilities were proven by two methods.

First, a comparison between systems with- and without graphical user interface was carried out. It was shown that the user-friendliness of systems with graphical user interface was judged to be significantly superior in all seven principles of ISO 9241 Part 10, to those systems without graphical user interface. The fact that the questionnaire ISONORM 9241/10 was able to clearly differentiate between software-systems is seen as first proof of the successful characterization of ISO 9241 Part 10. Moreover, it was shown that through ISONORM 9241/10 the user-friendliness of systems with graphical user interface as well as systems without graphical user interface is not quite satisfactory.

Second, in the forum of participative software-development the instrument ISONORM 9241/10 was used as a first analysis of weaknesses. In the case of the user-group meeting, the extraordinary quality was that ISO 9241/10 delivered an accepted criterion through which users and software-developers could discuss the product under consideration. The final result consisted of many concrete recommendations for the new software to be developed.

The questionnaire ISONORM 9241/10 does not claim to be an alternative to deeper analysis, such as error analysis through observation (Brodbeck, Zapf, Prümper & Frese, 1993 [2]; Prümper, Zapf, Brodbeck & Frese, 1992 [21]; Zapf, Brodbeck, Frese, Peters & Prümper, 1992 [28]; for an overview of evaluation methods see Dzida, Wiethoff & Arnold, 1993 [5]). Nevertheless, ISONORM 9241/10 can be a helpful alternative if a robust, economical software-evaluation instrument is sought, which is closely oriented towards international standards.

However, before its worth as a software evaluation instrument can truly be accepted, additional validity- and reliabilty-tests should be performed (see Kirakowski

& Corbett, 1990 [12]). These tests are currently being conducted and the preliminary results are quite promising.

The questionnaire ISONORM 9241/10 is available in both German (Prümper & Anft, 1993b) [19] and English (Prümper & Anft, 1993c) [20].

References

1. Altmann, A.: Direkte Manipulation: Empirische Befunde zum Einfluß der Benutzeroberfläche auf die Erlernbarkeit von Textsystemen. Zeitschrift für Arbeits- und Organisationspsychologie 31, 108-114 (1987)
2. Brodbeck, F.C., Zapf, D., Prümper, J., Frese, M.: Error handling in office work with computers: A field study. Journal of Occupational and Organizational Psychology (1993, in press)
3. DIN 66 234/8 Bildschirmarbeitsplätze: Grundsätze ergonomischer Dialoggestaltung. Berlin: Beuth 1988
4. Dzida, W., Itzfeldt, W.D.: User-perceived quality of interactive systems. IEEE Transactions on Software Engineering SE-4, 270-276 (1978)
5. Dzida, W., Wiethoff, M, Arnold, A.G.: ERGOguide - The quality assurance guide to ergonomic software. Delft University of Technology/German National Center of Computer Science 1993
6. Flanagan, J.G.: The critical incident technique. Psychological Bulletin 51, 327-358 (1954)
7. Frese, M., Sabini, J. (eds.): Goal-directed behavior: The concept of action in psychology. Hillsdale: Erlbaum 1985
8. Frese, M., Schulte-Göcking, H., Altmann, A.: Lernprozesse in Abhängigkeit von der Trainingsmethode, von Personenmerkmalen und von der Benutzeroberfläche (direkte Manipulation vs. konventionelle Interaktion). In: W. Schönpflug, M. Wittstock (eds.): Software-Ergonomie '87. Nützen Informationssysteme dem Benutzer? Stuttgart: Teubner 1987, pp. 377-386
9. Frese, M., Zapf, D.: Action as the core of work psychology: A German approach. In: M.D. Dunnette, L.M. Hough, H. Triandis (eds.): Handbook of industrial and organizational psychology (Vol. 4). Palo Alto, CA: Consulting Psychologists Press 1994, in press
10. Hacker, W., Volpert, W., von Cranach, M.: Cognitive and motivational aspects of action. Berlin: Deutscher Verlag der Wissenschaften 1982
11. ISO 9241/10: ISO 9241 - Ergonomic requirements for office work with visual display terminals (VDTs), Part 10 - Dialogue principles - First Draft International Standard, February 2, 1993
12. Kirakowski, J., Corbett, M.: Effective methodology for the study of HCI. Amsterdam: North Holland 1990
13. Nielsen, J.: Finding usability problems through heuristic evaluation. In: P. Bauersfeld, J. Bennett, G. Lynch (eds.): CHI'92 Conference Proceedings. New York: ACM press 1992, pp. 373-380

14. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: J.C. Chow, J. Whiteside (eds.): CHI'90 Conference Proceedings. New York: ACM press 1990, pp. 249-256
15. Prümper, J.: Benutzerorientierte, iterative Software-Entwicklung in der Praxis. In: W. Coy, P. Gorny, I. Kopp, C. Skarpelis (eds.): Menschengerechte Software als Wettbewerbsfaktor. Stuttgart: Teubner 1993a, pp. 630-647
16. Prümper, J.: GUIs sind noch weit entfernt von einer optimalen Beurteilung - ISO 9241 Teil 10 als Leitlinie für SW-Entwicklung und -Bewertung. Computerwoche 25, 46 (1993b)
17. Prümper, J.: Wie benutzerfreundlich ist Ihre Software? Börsenblatt 20, 24-26 (1993c)
18. Prümper, J., Anft, M.: Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. In: K.H. Rödiger (ed.): Software-Ergonomie '93 - Von der Benutzungsoberfläche zur Arbeitsgestaltung. Stuttgart: Teubner 1993a, pp. 145-156
19. Prümper, J., Anft, M.: ISONORM 9241/10 - Beurteilung von Software auf Grundlage der Internationalen Ergonomie-Norm ISO 9241/10 (Questionnaire) 1993b
20. Prümper, J., Anft, M.: ISONORM 9241/10 - Software Evaluation based upon the Principles of Ergonomic Requirements ISO 9241/10 (Questionnaire) 1993c
21. Prümper, J., Zapf, D., Brodbeck, F.C., Frese, M.: Some surprising differences between novice and expert errors in computerized office work. Behaviour & Information Technology 11, 319-328 (1992)
22. Rauterberg, M.: Läßt sich die Gebrauchstauglichkeit interaktiver Software messen? Und wenn ja, wie? Ergonomie & Informatik 16, 3-18 (1992a)
23. Rauterberg, M.: An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. Behaviour & Information Technology 11, 227-236 (1992b)
24. Rosson, M.B., Alpert, S.R.: The cognitive consequences of object-oriented design. Human-Computer Interaction 5, 345-379 (1990)
25. Shneiderman, B.: We can design better user interfaces: A review of human-computer interaction styles. Ergonomics 31, 699-710 (1988)
26. Téeni, D.: Direct manipulation as a source of cognitive feedback: A human-computer experiment with a judgement task. International Journal of Man-Machine Studies 33, 453-466 (1990)
27. Ulich, E., Rauterberg, M., Moll, T., Greutmann, T., Strohm, O.: Task orientation and user-oriented dialog design. International Journal of Human-Computer Interaction 3, 117-144 (1991)
28. Zapf, D., Brodbeck, F.C., Frese, M., Peters, M., Prümper, J.: Errors in working with office computers. A first validation of a taxonomy for observed errors in a field setting. International Journal of Human-Computer Interaction 4, 311-339 (1992)

Quelle:
Prümper, J. (1993). Software-Evaluation based upon ISO 9241 Part 10. In T. Grechenig & M. Tscheligi (Eds.), *Human Computer Interaction* (S. 255-265). Berlin Springer.