

EDM

This is an EDM carrier sheet

EDM

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

This is an EDM  
carrier sheet

*Please Return to  
HUSAT Library*

HM 1080

copy ①

EDM

This is an EDM carrier sheet

EDM

*HUSAT Memo Number 1080*

# ***Usability Evaluation Methods Within the Design Process***

*M. Maguire*

*March 1997*

*The HUSAT Research Institute  
The Elms  
Elms Grove  
Loughborough  
Leicestershire  
LE11 1RG  
Telephone: (01509) 611088  
Fax: (01509) 234651*

# USABILITY EVALUATION METHODS WITHIN THE DESIGN PROCESS

Martin Maguire  
HUSAT Research Institute  
The Elms, Elms Grove  
Loughborough, Leics LE11 1RG, UK

Tel: +44 1509 611088  
Fax: +44 1509 234651  
email: m.c.maguire@lut.ac.uk

This paper considers the importance of usability evaluation within the system design process and summarises the various evaluation methods that may be employed. Examples of practical problems that may be faced when performing usability evaluations are discussed.

## 1. THE IMPORTANCE OF USABILITY

Usability is now widely recognised as critical to the success of an interactive system or product (Shackel and Richardson, 1992; Eason, 1984; Whiteside et al, 1988; Nielsen, 1994). Many poorly designed and unusable systems exist which users find difficult to learn and complicated to operate. This results in systems which become under used, misused or even to fall into disuse as frustrated users return to their original working methods. The outcome is costly for the organisation using the system, and harmful to the reputation of the company who developed and supplied it.

The benefits of performing a usability evaluation can be summed up as follows:

- **Increased Efficiency.** A system designed to sound HCI principles and tailored to the preferred way of working, will allow the user to operate effectively and efficiently rather than lose vital time struggling with a poorly designed user interface.
- **Improved Productivity.** A good interface will allow the user to concentrate on the task rather than the tool which if designed inappropriately can extend, rather than reduce the time to do a task as well as directly affecting other aspects of performance or quality.
- **Reduced Errors.** A significant proportion of so-called 'human error' can be attributable to a poorly designed user interface. Avoiding inconsistencies, ambiguities or other interface design faults will reduce user error.
- **Reduced Training.** A poorly designed user interface can prove a barrier to an otherwise technically sound system. A well designed interface can reinforce learning, thus reducing training time and effort.
- **Improved Acceptance.** This is not usually the primary objective in improving the interface, but is nevertheless an important outcome. Most

users would rather use and would be more likely to trust a well designed interface which provides the information in a format which is easy to assimilate and use.

## 2. BASIC APPROACH TO USABILITY EVALUATION

### 2.1 Understanding the users and tasks

The starting point for performing a usability evaluation is to gain a thorough understanding of the requirements of the user population. It is important to consider variations in requirements between different types of user.

A useful approach is to perform a 'stakeholder analysis' whereby all those people with a 'stake' or interest in the system are identified and their particular needs specified. For example, a public information system situated in a local library or advice bureau, where the primary users will be the general public. However, while the system may be designed to be easily used by the public, it may alternatively be used on their behalf by a local, professional adviser. A third user group will be the staff who enter the information into the system database and maintain it on a regular basis.

All these users will want to use the system in different ways and will have different skills, abilities and levels of technical knowledge. This variance must be taken into account if the system is to meet the differing needs. For a member of the general public who is using a system for the first time, a 'usable' system will be one which guides them step by step through the stages of the task (gaining information, checking their bank balance, booking a theatre ticket and so forth). In contrast, an information adviser answering enquiries on behalf of members of the public, perhaps over the telephone, will require the system to allow rapid access to the information they require.

A range of tools and techniques exist for investigating the requirements of the user population. One of the outcomes of the Esprit HUFIT project, Human Factors in Information Technology, (Allison et al, 1992) was a set of tools for identifying user types, their needs and characteristics, and translating this into functional requirements. A second tool is 'Usability Context Analysis' (Maissel et al, 1991), a paper based questionnaire and guide for establishing an agreed detailed description of who uses the software to be evaluated, the tasks that it is to be used for, and the environment in which it is used. This was developed as part of the ESPRIT MUSiC project (see below).

Studies of the user population will result in an understanding of the user requirements of the system, and the usability goals that it should achieve and be tested as part of the evaluation process. These may include the following:

- Effectiveness — allowing the user to perform their required tasks fully.
- Efficiency — allowing the tasks to be performed quickly and easily.
- Intuitiveness — allowing the system to be learned within a reasonable time.
- Supportiveness — assisting the user throughout the interaction and helping them to overcome problems that may occur.

## 2.2 Prototyping the user interface

In carrying out evaluations of a user interface design, a valuable technique is the use of computer simulations or rapid prototypes of the future system. This may consist of a series of screen layouts and a partial database allowing potential users to interact with, visualise and comment on the system. Such simulations or prototypes can be produced both quickly and easily in the early stages of the system development cycle for evaluation by Human factors experts, user representatives and members of the design team.

Changes to the design may then be made rapidly in response to user feedback so that major problems with the design can be identified before system development begins. This helps to avoid the costly process of correcting design faults in the later stages of the development cycle.

## 2.3 Evaluation with representative users

Having developed a system simulation or prototype, it should be tested with representative end users. Careful consideration must be given to the future context in which the system will be used and to defining a realistic plan for measuring the usability of the system.

This involves:

- Selecting a sample of representatives of user groups who will eventually use the system
- Defining realistic representative tasks to perform
- Setting up the test environment to reflect working conditions.
- Choosing a suitable evaluation method and set of measures to collect.

One approach may be to consider user performance with a system (that is, how long does it take to perform a set of tasks, how many problems arise and so forth). Another may be to record users' attitudes and feelings towards a system having used it. Typically a combination of such approaches are used. This will identify, for example, the situation where the user performs well but does not enjoy using the system. Furthermore, measurements of mental workload (using questionnaires or physiological recording) may show where users are overloaded or highly stressed despite producing a good performance level.

Many IT organisations such as IBM, Hewlett Packard, DEC and ICL have invested in advanced, dedicated laboratories for performing such usability evaluation work. This facility may consist of a user area which can be set up to reflect a range of operational contexts and a control area for observation by the human factors evaluator. A one-way mirror may separate the two areas so that the user can be observed by the evaluator in the control area although the evaluator cannot be seen by the user.

An evaluation of a prototype system will typically involve running test sessions with between 10 to 25 users. User interactions and comments can be recorded during each test session on videotape for later analysis. The output of a usability evaluation is normally a report describing the process of testing that was

carried out, the results obtained, and recommendations for system improvement. An additional and useful technique is to create a short film composed of video clips from the user sessions illustrating key problems that were encountered with the prototype or facilities that work especially well. This provides a means of emphasising the results of the evaluation to the design team. The results may also be passed on to other departments such as Marketing or Senior Management to support the case for the development of a new product or innovative set of features.

While usability evaluations require care in their planning and performance, in practice they often need to be carried out within a short timescale as part of an iterative development cycle where prototypes are provided by a commercial client organisation and changes made to it as a basis for further user testing.

### 3. MUSIC EVALUATION TOOLSET

A growing area is the development of standard methods and measures for performing usability evaluations. HUSAT has been involved with the Esprit 5429 MUSIC project (Measuring Usability in Context) which has developed a set of standard tools and techniques for measuring software Quality of Use or usability (Bevan and Macleod, 1994). These tools incorporate a set of clearly defined methods and metrics for investigating different aspects of usability. The tools also support the diagnosis of problems underlying the metrics obtained during an evaluation. The user-based testing tools include:

- Usability Context Analysis - a paper based questionnaire and guide for establishing an agreed detailed description of who uses the software to be evaluated, the tasks that it is to be used for, and the environment in which it is used.
- Performance measurement tools - a method and associated software to guide the process of evaluation of user performance with a system. This includes defining a series of representative tasks (using the Usability Context Analysis approach), capturing the user sessions on videotape and obtaining measurements across the user sample. These include user effectiveness, user efficiency, relative user efficiency (compared to an expert), and productive period (productive time not spent in overcoming problems and seeking help). See Rengger et al, 1993.
- Measurement of user perceived quality - this is accomplished using SUMI, a questionnaire based method for assessing the software user's view on the scales of Learnability, Efficiency, Affect (likeability), Control and Helpfulness, (Kirakowski and Corbett, 1993).
- Measurement of mental effort - the use of questionnaires to provide an indication of the mental load the user is under during their software sessions. This is a complementary method to the user performance and perceived quality methods (Arnold et al, 1994).

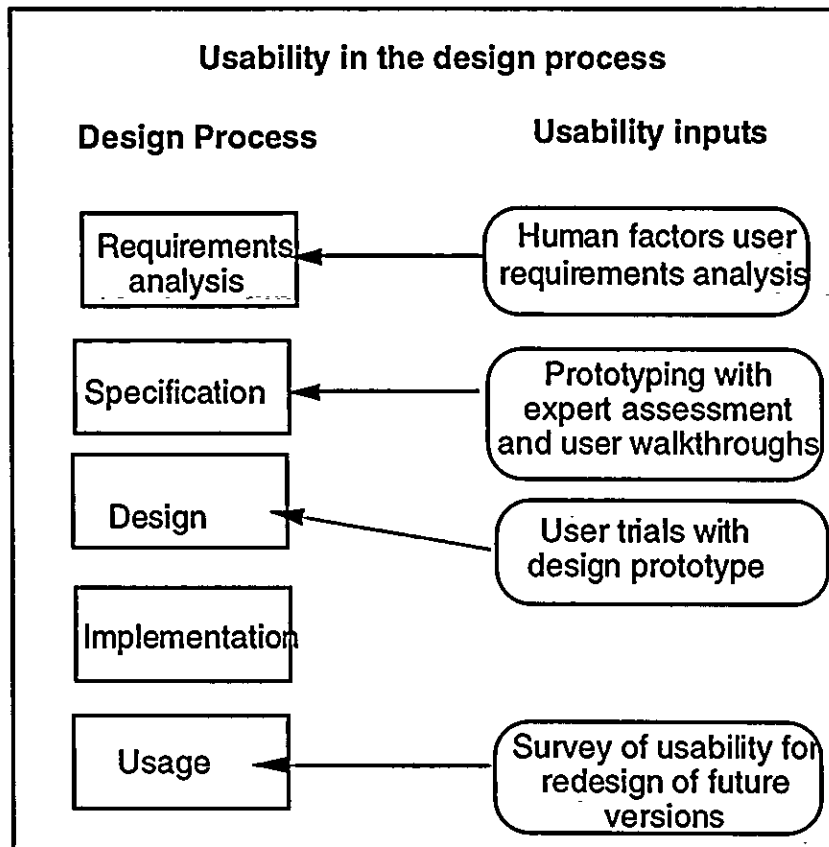
Within the system lifecycle, it may be desirable to obtain clear unambiguous measures giving an indication of the usability of the current version of the system. Such measures can be used to compare different system versions, or to compare the system with a competitive product.

The MUSiC toolset thus has been developed to have relevance throughout the system design cycle. At the early specification stage, analytic metrics may be used to assess possible problems before a system prototype is built (Gunsthövel and Bösser, 1991). At the prototyping stage, when the user can interact with a prototype, then the user performance, user perceived quality and mental effort tools can be applied to help improve the design. When the system has been released, similar methods can be applied as a basis for making changes at the alpha and beta testing stages. After full release, evaluation measures may be taken in preparation for building a new version of the system.

MUSiC tools have been applied successfully in a range of different application areas (e.g. management information systems, fleet management, and systems for stock market trading) and has been taken up by a number of large companies. However they do require a certain amount of effort to be invested in them if the formal evaluation process is to be carried out successfully and valid results obtained.

#### 4. METHODS OF EVALUATION

An evaluation can be carried out in a number of ways. The approach selected will depend on the stage of system development (such as an early prototype or operational system), the availability of a representative sample of users, and time and resources available.



##### 4.1 Expert assessment

This is one of the most straightforward ways of identifying problems related to the system. The human factors expert will start by gaining a thorough understanding of the user characteristics and the nature of the task and working environment in discussion with the design team and desirably with user representatives. The expert will then study a prototype or demonstration of the system and mentally pose a number of questions which will highlight problems of the intended system and lead to recommendations for improving it. At the same time the expert may identify problems instinctively i.e. where some feature contrasts with the experts view of good practice.

The main advantage of an expert appraisal is that is a quick and easy way to obtain feedback and recommendations or ideas can easily be obtained. The disadvantages are that experts may have personal biases towards specific designs and it is often hard to set aside one's expertise and assume the role of the user.



In order to make the appraisal, the expert is likely to ask the following types of question:

- (i) If the user wants to do something simple, is it simple to do?
- (ii) Considering a series of task situations, how does the system fit into each of them?
- (iii) How will the system fit into what really happens in the working environment e.g. if the controller requires that alarms must be acknowledged, will this cause annoyance and frustration?
- (iv) How would the typical user react to seeing this screen display ? or using this means of inputting values?

The expert will be able to draw upon effective implementations of similar systems he or she has encountered in the past as well as applying general principles such as:-

- Simplicity
- Consistency
- Provide shortcuts for Experienced Users
- Make it difficult to make an error, but easy to correct one.
- Provide useful feedback and prompts
- Reduce memory load
- Keep the user in control

The expert will also be able to assess a prototype user interface to check its conformance to relevant legislation and standards. As of the 1st January 1993 user interface software produced in the UK is governed by legislation via the Display Screen Equipment regulations, the premise being that poor user interfaces can lead to sub-optimal performance and stress. It may be argued that compensation claims, now currently targetted at physical aspects of the workstation, will increasingly concentrate on negligently produced user interface software. Two aspects are currently important. The European Directive 90/270/EEC, (CEC 1990) on Display Screen Equipment requires that:

- software must be suitable for the task
- software must be easy to interpret
- principles of software ergonomics must be applied.

ISO 9241 is a multi-part international standard, in the process of development, containing 'usability statements' (ISO, 1993). British Telecom are amongst companies now requiring compliance with these standards on the part of their suppliers although it is desirable to aim for the optimal design of software rather than merely comply with minimum standards.

## 4.2 User walkthroughs

A related approach to expert appraisal of a system is a 'walkthrough' of the different parts of the system with users, a Human Factors expert and possibly designers present to discuss possible problems and to recommend redesign

(Lewis and Polson, 1992). The approach is to identify mismatches between what the system provides and what the users wish to do in a structured way. One person, normally the Human Factors expert will chair the activity, writing down questions, answers and possible solutions as the discussion progresses. That person also takes responsibility for summarising the outcomes of meetings and for ensuring that the discussion stays 'on track'. Depending on the purpose of the exercise, a typical user task or set of tasks that the system is intended to support is selected to walk (or talk) through. The team then works through this task in great detail considering all user actions one by one, in the sequence they would occur in the task under consideration. The effect of these actions on the system, and changes in system states are considered as well. All along the way, the team imagines a range of scenarios 'what if the user does X right now?' in an effort to think through the task and its implications. All possible stumbling blocks are identified and different solutions applied to overcome them. The Human Factors expert then analyse the discussion and consult with the users to clarify comments and the technical staff to determine what changes to the system are possible before recommendations are made.

While a walkthrough may be a highly structured process it has been argued that can be restrictive and encourages users to discuss minute aspects of the user interface and ignoring alternative design possibilities. Lindegaard (1994) argues for a more relaxed method whereby the discussion group adopts a more relaxed attitude and the walkthrough can thus be seen as a method for collaborative design.

The advantage of a user walkthrough and discussion is that it is cheap and relatively easy to administer. Time can be saved when compared with demonstrations and interviews with users on an individual basis. This type of situation is also particularly good for obtaining ideas and agreement on recommendations. Negative aspects are that some users may be inhibited in group situations and one user can dominate the session. Sessions are also difficult to structure and, where the subjects are not familiar with the goals of the evaluation, issues may become confused and disorganised.

A great deal of valuable information can be collected in a short period of time, but the technique demands skilled administration to ensure that all opinions are represented and not simply a consensus based on 'middle ground'. For this reason it may work best when the user group is composed of a homogeneous sample of users e.g. secretaries, rather than a group containing, say, supervisory and shop-floor users.

#### 4.3 User trials

The most revealing method of usability evaluation is to set up trials with the system whereby a representative set of users are asked to perform a series of tasks. This may be set up in a controlled laboratory environment, at the site of the development organisation or in the field within a customer organisation. The aim would be to gather information about the user's performance with the system, their comments as they operate it, their post test reactions and

observations made by the evaluator. The main benefit of this approach is that the system will be tested under conditions close to those that will exist when it is used 'for real'. While technical designers and human factors experts may diagnose a large proportion of potential system problems, experience has shown that working with users will reveal new insights that will affect the system design.

Data from user trials can be captured in a number of ways:

- Automatic system monitoring may be set up whereby the system itself records interaction events of importance. These can be timestamped to provide accurate information about the user's performance or their methods of navigating through the system. This method is good for ensuring that all events are recorded and accurately timed. The data may also be used to replay the user interaction if necessary. However the capture of large amounts of low level data is difficult to interpret in terms of usability. Where possible, the capture should record higher level data events e.g. entering a new record, deleting a record, or access to help facilities etc.
- Observation and manual recording by the evaluator of events during the interaction session such as: time to complete task, points of apparent user difficulty, errors made, general demeanour of the user, approach to using the system etc. While this method is very demanding, it means that useful data is recorded immediately on paper from which results can be obtained straightaway. The data also represents first hand feedback of user interaction covering aspects that automatic monitoring omits such as the use of documentation, user behaviour etc. However the capture of large amounts of low level data is direct observation and note taking may be seen as intrusive by some users if care is not taken to put them at their ease during the test session.
- A third method is to record the user session onto videotape. This technique has had a major impact on the process of studying human-computer interaction since a comprehensive record of a session can be recorded and then analysed at leisure to gather both performance data, user behaviour and verbal comments during the interactions. A number of evaluators can analyse the same recording thus increasing the reliability of data analysis. Also parts of the video can be edited into a short film to show design staff at first hand problems that users the experienced. However the analysis of video tapes can be very time consuming (even with special video analysis software) and can become a heavy burden if a large build up of unanalysed tapes occurs.
- Having interacted with a system, user's can be asked to comment upon it and to give opinions and ratings using a number of scales (such as its ease of use, flexibility, helpfulness etc.) These ratings can be combined to provide numerical data reflecting the users opinions of the system. User suggestions for improving the system can also be generated at the same

time. This method is nearly always appropriate to use as it is a quick and easy way of obtaining user feedback. However questions and the wording for rating scales needs careful design if the users are to understand them clearly and thus give accurate answers.

The results of user trials can be combined to provide both objective data on the usability of the system, descriptive data from the users responses, and recommendations on how the system may be improved. It is also important to highlight successful features of the system that operated well under user scrutiny.

#### 4.4 User surveys

For systems already in use at a number of sites, a questionnaire survey can be an effective method of gathering information on the system's usability from a wide range of users. Such surveys can be carried out over the phone, by post, or by visiting sites running the system and gathering user reactions at first hand. On site interviews can be combined with observation of the system in use. A survey may be based on a structured questionnaire or interview schedule so that comparable data is obtained across the user population. Also data can be gathered based on rating scales so that usability scores relating to people's opinions can be generated, possibly to compare two systems or an older and newer version of the same system.

The pitfalls to avoid in performing interviews are as follows:

- not making the aim of the interview clear to the interviewee
- asking questions that the interviewee does not know the answers to, thus forcing them to make guesses
- leading the interviewee or not being neutral, thus biasing the answers given
- asking too many questions

Pitfalls for questionnaires include:

- written questions and the method of answering them being unclear
- poorly structured questionnaire causing respondents to miss out parts
- the use of too many open questions, the answers being hard to analyse.

### 5. EXAMPLES OF EVALUATION WORK

#### 5.1 Evaluation of home products

HUSAT have participated in a European project collaborating with major companies manufacturing consumer electronics. One activity involved evaluating simulations of user interfaces to home products with members of the public in an evaluation laboratory. The companies produced initial versions for Human Factors experts to comment on. This feedback was incorporated into revised simulations which were then tested fully with members of the public performing tasks such as programming a video recorder or setting up a heating control programme. During the trials, data was captured of user performance (for comparison with previous versions) and subjective comments from the

users themselves. The sessions were recorded on video tape for more detailed analysis. The written reports to the companies, containing both evaluation data and recommendations for improvement, were supplemented with video clips showing key problems experienced by users as well as highlighting facilities that worked well. During the project seven rounds of evaluation were conducted as part of an iterative design cycle. This proved to be an effective means of developing the user interface designs.

Practical problems that needed to be addressed were mainly due to time constraints on test sessions. Users were asked to look at a device simulation on screen, given a brief introduction to it, and then asked them to complete a series of tasks. Clearly this type of testing is addressing the learnability of the device but not its day to day usage after a period of time when users will have explored a range of facilities and either continued to use them or given up. Also the learning process in the home may be assisted by others or an instruction book. While the project wished to study the intuitiveness of the products, the results have to be interpreted in this context of testing.

## 5.2 Evaluation of financial systems

HUSAT have performed a number of usability evaluations of a range of systems for money-market traders and dealers, for a company which recognises the need to improve the usability of its systems. These evaluations have been conducted at the client's site allowing the client to employ representative users from within the local area (the city of London). Again, due to the need for rapid feedback to design staff, these evaluations have been conducted quickly — typically a week of preparation, a week of testing and a week to analyse the data and report results. Following the successful application of this evaluation process to a small number of key products being developed, it is now company policy that all existing and future products will be evaluated with representative end users.

The types of problem that were encountered were that the co-operation of users could be difficult to obtain. Due to the nature of their job, some users (e.g. stock market traders) often had to cancel their test sessions. This often meant that gaps in the testing schedule arose and could only be filled by less representative personnel.

## 5.3 Comparison of hardware

A full set of usability trials were conducted on two automatic banking machines focussing on their ease of loading with money and receipt paper. One was an existing fully operational machine, while the other was a prototype mocked up prior to its manufacture. Representative users (e.g. bank staff personnel) performed a range of tasks to investigate the accessibility and ease of use of the two machines and, where appropriate, comparisons were drawn between them. These trials were supplemented by an expert analysis of the machine on anthropometric grounds. The evaluation was able to highlight which parts of the new design worked well and which needed further attention before being committed to being developed for full production.

While the prototype device was found excellent for testing issues of accessibility, this was less realistic for ease of operation where it was found difficult to compare paper loading on the prototype with the same operation on the fully functional machine. Similarly the absence of wiring made the prototype more attractive than the operational machine and so it was felt that for future testing the shielding of electronics would be helpful for making the two devices comparable. The testing also showed that some tasks were not always necessary and so the results for them may not be particularly relevant.

## 6. CONCLUSION

The achievement of usability in systems design requires a combination of:

- Articulating, specifying and understanding user requirements.
- Human centred, application driven iterative user interface design.
- Appropriate usability evaluation, ensuring that:
  - subjects are representative users
  - realistic tasks to be performed are designed
  - the test environment reflects the working conditions
  - appropriate measures and methods of measurement are taken.

It is becoming increasingly obvious that a high level of usability is the key to future commercial success with the myriad of systems, personal computers and computer services, telecommunications products and services.

To ensure a successful outcome, the design team must satisfy the needs of the user when the development is complete. To do that, the users of future systems must be represented throughout the process as the best option for successful products in the future.

## ACKNOWLEDGEMENT

HUSAT's work on the MUSiC project was funded by the Commission of the European Communities.

## REFERENCES

Allison G., Catterall, B., Dowd, M., Galer, M., Maguire, M., and Taylor, B. (1992) Chapter 2: 'Human factors tools for designers of IT products'. In: Methods and tools in user-centered design for IT, Eds: M. Galer, S. Harker and J. Ziegler, North Holland, Human Factors in Information Technology series, no. 9.

Arnold, B., Houwing, E., Wiethoff, M. (1994)  
Work carried out at Technische Universiteit Delft, Kanaalweg 2B, Postbus 5050, 2600 GB delft, The Netherlands.

Bevan, N., and Macleod, M. (1994)  
'Usability measurement in context', Behaviour and Information Technology, vol. 13, nos. 1 and 2, pp 132-145.

CEC (1990)

'Minimum safety and health requirements for work with display screen equipment directive (90/270/EEC)', Official Journal of the European Communities No L 156, 21/6/90.

Eason, K. D. (1984)

'Towards the experimental study of usability', Behaviour and Information Technology, vol. 3, no 2, pp 133 - 143

Gunsthövel, D. and Bösner, T., 1991

'Predictive metrics for usability', Proceedings of the 4th International conference on human computer interaction, Amsterdam, Elsevier. ed: H-J. Bullinger.

ISO (1993)

'ISO CD 9241-11: Guidelines for specifying and measuring usability'. Part 11 of 17 part standard: 'Ergonomic requirements for office work with visual display terminals'.

Kirakowski, J. and Corbett, M. (1993)

'SUMI: the software measurement inventory', British Journal of Educational Technology, vol 24, pp 210-212.

Lewis, C.H. and Polson, P.G. (1992)

'Cognitive walkthroughs: A method for theory-based evaluation of user interfaces', CHI'92 Striking a Balance, Tutorial, 4th May.

Lindgaard, G. (1994)

'Usability testing and system evaluation - a guide for designing useful computer systems', Chapman and Hall Computing Series. ISBN 0-412-46100-5.

Maissel, J., Dillon, A., Maguire, M. Rengger R. and Sweeney, M. (1991)

'Context guidelines handbook', MUSiC project deliverable IF2.2.2 (HUSAT Research Institute, Loughborough, Leics, UK and National Physical Laboratory, Teddington, UK).

Nielsen, J. (1994)

Behaviour and Information Technology, Special issue: usability laboratories, Vol. 13, No. 2, Taylor and Francis, London

Rengger, R., Macleod, M., Bowden, R., Blaney, M. and Bevan, N. (1993)

MUSiC performance measurement handbook (National Physical Laboratory, DITC, Teddington, UK)

Shackel, B and Richardson, S. (1991)

Human Factors for Informatics Usability, Cambridge University Press. ISBN 0-521-36570-8.

Whiteside, J., Bennett, J. and Hotsblatt (1988)  
Usability Engineering: our experience and evolution. In: Helander, M (Ed)  
Handbook of Human Computer Interaction, pp 791 - 817, Elsevier, Amsterdam