

1 Activities to be completed with user participants

Process overview, system under test overview and setting information

Test session checklist (completed by the test administrator)

- Greet user and introduce development observers, data loggers and others present in the room.
- Explain the usability lab and show video camera views and such until the user is comfortable with the equipment.
- Proceed to the test area.
- Explain hardware setup to user and confirm that the user is comfortable with it.
- Seat user and sit close by to explain the test session.
- Have user complete the respondent profile questionnaire.
- Have user sign the video/audio tape and disclosure agreement.
- Ask the user if they have any questions.
- Have the user read the overview sheet titled "The Process".
- Review participant package (task scenario booklet, help documentation, XMM overview and setting descriptions).
- Ask the user if they have any questions.
- Have the user read the sheets titled "XMM overview" and "The Setting".
- Ask the user if they have any questions.
- Review with the user the help resources available for example User's Guide and on-line help.
- If user is a novice suggest the Windows tutorial to become familiar with common Windows processing techniques.
- Ask the user if they have any questions.
- Review and demonstrate the "thinking aloud" process and complete practice exercise until the user is comfortable.
- Ask the user if they have any questions.
- Ask the user to wait until instructed to open the task scenario booklet.
- Return to lab area, verify logging is ready to begin and start video tape.

The Usability Process

Usability evaluations seek to determine if the people who use the product can do so quickly and easily to accomplish their own tasks. Usability applies to every aspect of the product in which a person interacts, such as hardware, software, menus, icons, messages, documentation, and help. Evaluations are designed to solicit feedback from participants, focusing on areas of concern identified by our customers. An evaluation typically involves several participants, each of whom represents a typical user.

Once all evaluation sessions are completed, we compile the feedback received from each participant, along with our notes. We then prepare a final analysis report and a highlight tape, which we present to the customer for review.

Our analysis is not a mandate for changes in the product. Our intent is to provide a base of information from which our customers can make decisions to resolve problem areas in the interface and the work process. We also note positive feedback from participants, to indicate those areas in which the design is successful.

Do you have any questions?

Xerox Management Model Overview

The Xerox Management Model allows individuals to integrate the quality strategy into how the business is run. The model provides the framework for combining all the elements needed for achieving business results and continuous improvement.

The software allows you to access the fundamental information and resources you will need to understand the model and find related documents and tools you will use on the job every day.

Do you have any questions?

The Setting

You are in your office preparing for your day. XMM is already installed on your computer. You use the XMM software to help you organize your data and to provide a framework for your daily activities.

Do you have any questions?



INTERNAL MEMO

TO

Distribution

FROM

Denise D. Pieratti
Mgr. Usability & Technical Doc.
GP&IM/GSD/OASIS
810—01A
8*223—9448
D2:WBST845E:Xerox

SUBJECT

Looking for usability evaluation
participants

DATE

April 17, 1995

An exciting opportunity awaits you! The Usability and Technical Documentation group, part of GP&IM, is looking for people to help evaluate, and review new applications being developed to support Xerox business process re-engineering efforts. As a participant in a software evaluation, you will have an opportunity to influence the direction, appearance, and “look and feel” of the software applications being developed to support the entire Xerox community.

We need to identify potential candidates for usability evaluations that represent the diversity of the Xerox. We are looking for people that have these characteristics:

- No computer experience all the way to “power users”
- Lots of management experience, to those with no management experience
- Exempt employees, as well as non-exempts and contracts
- A diversity of educational experience and levels

In other words, we are looking for people who work for Xerox. Will you help us?

If you would like to participate in usability evaluations of new software applications, please complete and return the survey form. When your profile matches the profile of the intended users of the software being evaluated, we will contact you and schedule a usability evaluation at your convenience.

A software evaluation generally takes no more than two hours, but you will be told all the details when we contact you. If you are interested in joining our bank of potential usability evaluators, please complete the following survey and return it to the address on the form.

We hope that you will take an active role in shaping the software being developed in support of Xerox 2000! Your participation is greatly appreciated!

1 Test Day Schedule of Daily Events

Monday March 6th

1:00 - 3:00 P.M Overview presentation scheduled

Tuesday March 7th

7:30 A.M. Arrive at site
Power up lab and do prep testing

9:00 A.M. **First** test subject arrives
Run through test session checklist

9:15 A.M. Begin test

10:30 A.M. End test
Debrief user and PostTest questionnaire filled out

11:00 A.M. **Second** test subject arrives
Run through test session checklist

11:15 A.M. Begin test

1:00 P.M. End test
Debrief user and PostTest questionnaire filled out

1:15 - 2:00 P.M. LUNCH

2:00 P.M. **Third** test subject arrives
Run through test session checklist

2:15 P.M. Begin test

3:30 P.M. End test
Debrief user and PostTest questionnaire filled out

4:00 P.M. Compile notes & identify any outstanding problems to communicate
immediately
Check & supply new batteries for test equipment
Label all test materials and store
Prepare test materials for next test day

Test Day Schedule of Daily Events

Wednesday March 8th

- 7:30 A.M. Arrive at site
Power up lab and do prep testing
- 9:00 A.M. **Forth** test subject arrives
Run through test session checklist
- 9:15 A.M. Begin test
- 10:30 A.M. End test
Debrief user and PostTest questionnaire filled out
- 11:00 A.M. **Fifth** test subject arrives
Run through test session checklist
- 11:15 A.M. Begin test
- 1:00 P.M. End test
Debrief user and PostTest questionnaire filled out
- 1:15 - 2:00 P.M. LUNCH
- 2:00 P.M. **Sixth** test subject arrives
Run through test session checklist
- 2:15 P.M. Begin test
- 3:30 P.M. End test
Debrief user and PostTest questionnaire filled out
- 4:00 P.M. Compile notes & identify any outstanding problems to communicate immediately
Check & supply new batteries for test equipment
Label all test materials and store
Pack up lab for transportation



Understanding Your Participation

Please read this page carefully.

Xerox Corporation is asking you to participate in evaluating a new product. By participating in this evaluation, you will help us improve this and other Xerox products.

We will observe you and record information about how you work with the product. We may also ask you to fill out questionnaires and answer interview questions.

We will videotape all or some of the interview and your work. By signing this form, you give your permission to Xerox to use your voice, verbal statements, and videotaped pictures for the purposes of evaluating the product and showing the results of these evaluations. We will not use your full name.

You will be working with a product that is still being developed. Any information you acquire about this product is confidential and proprietary and is being disclosed to you only so that you can participate in the evaluation. By signing this form, you agree not to talk about this product to anyone. You may tell them that you helped to evaluate new software.

If you need a break, just tell us.

You may withdraw from this evaluation at any time.

If you have any questions, you may ask now or at any time.

If you agree with these terms, please indicate your agreement by signing here:

Please print your name _____

Signature _____

Date _____

What type of applications do you use?

Word Processing (like Word)

Spreadsheet (like Excel)

Database (like Access)

Presentation (like PowerPoint)

Project Management (like MS Project)

Desktop Publishing (like PageMaker)

CD ROM/Multimedia

Other _____

What do you typically use your computer for?

Games and Pleasure

Graphics

Accounting/Finance

Data storage (i.e., data bases)

Word Processing

Decision Support

Programming

Other _____

Please return this questionnaire to Usability & Technical Documentation at mailstop: 810-01A by
4/26/95.

Methods for successful “Thinking Out Loud” procedures

Developed by Judy Ramey, University of Washington, with additions by Usability Analysis & Design, Xerox Corporation

1 In General

- When you are screening the participants for a study, notice how they respond to your questions. Decide on a strategy for engaging the participant before they arrive for the usability study.
- Be careful of the social dynamics you set up with the participant
 - Don't joke, indulge in sarcasm, flirt, or betray your own nervousness
 - Maintain a professional, neutral persona
 - Keep yourself “small” in relationship to the participant. Sit slightly back from the participant, in a chair that is lower.
 - Avoid wearing heavy perfume or aftershave. The participant may have allergies to the odor or find it distracting.
 - Don't wear suggestive, revealing, or tight, uncomfortable clothes.
- Don't bias the participant
 - Don't betray your own views or opinions of either the participant's level of skill
 - Don't let the participant become aware of any bias you may have about the product.
- Avoid interactions with the user that can shift the focus from the user's domain to the designer's
 - Don't expect the user to tell you how to fix problems
 - Don't expect the user to answer other design questions
 - Always keep the focus of attention on the user, not yourself. Avoid “I” statements and long explanations of how the system works.
 - Stay in the relationship with the participant. Don't worry about the next question you are going to ask.
 - Write down design ideas so that you don't need to worry about forgetting them after the test.

- Don't let yourself get impatient!

When the participant seems to have a problem, they can often unravel it without your help.

When you feel you should jump in, count to ten first.

If you jump in too soon, you lose valuable data and they become dependent on your help.

- Learn to probe in a neutral way to get information on which to base your design improvements

2 Techniques that encourage thinking out loud

- Prompting
- Echoing
- "Conversational disequilibrium"
- Summarizing at key junctions

3 Prompting

- Focus on tasks, not features

Don't ask "Do you like that dialog box?" but "Did that dialog box help you reach your goal?"

- Focus on questions, not answers
- Explore user thinking in a neutral way

Don't be too quick to assume that the user is lost or having a problem.

Don't say, "What is your problem here?," but ask, "What is your goal?" or "What are you thinking you should do here?"

- Don't betray your own interests or point of view by your comments, emphasis, "waking up" and getting interested, showing in facial expression or vocal tones that you disagree
- Good user-focused questions:

What is your goal?

What did you expect when you did that?

How did you expect that to work?

Can you tell me what you were thinking?

What do you want to accomplish here?

Describe the steps you are going through here.

How did you feel about that process?

Tell me about your thinking here.

What did you expect to happen when you . . . ?

4 Echoing

- Repeat their own word or phrase back to them as a question: “That message is confusing?”

Echoing sets up a social dialog and reinforces social conversation expectations: they say something, you repeat it, they say the next thing because that is what is expected in conversation.

- Don't put words in their mouth, or offer interpretations

If they say, “I'm not sure what to do here,” don't say, “So you are confused because the menu bar is unclear?”

If they say, “That didn't happen like I expected, don't ask, “So you thought that the task menu would be displayed here?”

- Signal that you are listening (Mmm hmm . . .)

5 “Conversational disequilibrium”

- Let your statements trail off and end in an upswing, as if you were asking a question. The participant will usually complete your statement.

“And you were expecting. . .?”

“And your goal is . . .?”

- Signal that you are there, you are interested, but that it is still their turn to talk (mmm hmhhh)
- Speak softly

6 Summarizing at key junctions

- When you have learned something new that is key to understanding, summarize the event and the thinking that the user explored, very briefly. Users may offer more detail about their thought process.
- Keep the recorder on or keep taking notes after you think that the test session is finished. Users will often make interesting reflections about their processes during the casual remarks at the end of the session.

Thank You!

We appreciate the time you have taken to participate in the evaluation of the Xerox Management Model. Please except these tickets as an expression of our gratitude.

Your participation in the XMM evaluation assists in the development of an effective and usable tool. The comments and observations you expressed during this evaluation will be reviewed by the development team and included in our final report.

If you have any questions or we may be of assistance to you, please feel free to contact us.

*Wendy W. Naughton
Usability Specialist
(716)383-7250
wendy_naughton@mc.xerox.com*

*Denise D. Pieratti
Manager, Usability & Technical Documentation
(716)383-9448
D2@mc.xerox.com*

1 Usability Severity Codes

Severity Code	Description
1 – Severe (1=Showstopper)	<p>An emergency condition that causes the customer’s system to fail or causes customer data to be lost or destroyed. A showstopper usability bug can also be one that is likely to cause frequent data integrity errors. There is no workaround to these problems. A key feature needed by many customers is not in the system.</p> <p>EXAMPLES:</p> <ul style="list-style-type: none"> • An obvious competitive feature is missing, or the customer expects something reasonable, but the feature isn’t there. • Critical data is lost often because users are expected to “know” to save their data or do something that is complex to preserve the data. For example, if the application allows users to exit without prompting them to save their changes or asking if they want to save a session. • The usability problem is likely to result in an error that will cost the customer a large amount of wasted time or money (for example, a problem that would delay the recovery of a transaction database) • The usability problem is likely to lead to a significant safety or health risk. • Legal issues like missing copyright notices or improper use of another company’s trademarks
2 – High	<p>A serious condition that impairs the operation, or continued operation, of one or more product functions and cannot be easily circumvented or avoided. The software does not prevent the user from making a serious mistake. The usability problem is frequent, persistent, and affects many users. There is a serious violation of standards.</p> <p>EXAMPLES:</p> <ul style="list-style-type: none"> • There is no keyboard access method for a feature. • The way to work around the problem is very hard to explain or hard to remember. The problem is likely to result in many support calls. • Missing or undocumented parameters, menu commands, or other important information in the interface or documentation. • Serious violations of corporate user interface design standards • Offensive, inappropriate, or non-standard icons for standard functions • Missing feedback for critical operations
3 – Medium	<p>A non-critical, limited problem (no data lost or system failure). It does not hinder operation and can be temporarily circumvented or avoided. The problem causes users moderate confusion or irritation.</p> <p>EXAMPLES:</p> <ul style="list-style-type: none"> • Keyboard access is possible, but not easy. Tab order does not follow the user interface guidelines • A common menu item or toolbar button does not do what is normally expected. • Incorrect or no error messages (this could be High or even a showstopper in some cases if a bad error message results in lost data) • Incorrect data are generated (this could be a high or a showstopper bug if it causes the user to make a bad decision that has significant implications for the user or company) • Noticeable performance inefficiencies (this can be a high if the performance is bad enough) • Use of the wrong control in the interface (the wrong control causes some moderate performance degradation) • Missing or incorrect titles or units • Unneeded clutter that does not have a major impact on user performance
4 – Low	<p>Non-critical problems or general questions about the product. There are minor inconsistencies that cause hesitation or small aesthetic issues like labels and fields that are not aligned properly.</p> <p>EXAMPLES</p> <ul style="list-style-type: none"> • Typos (this could be a higher priority if they are in key places like menus or splash screens) • A menu item does not use the recommended access character • Small cosmetic problems

SUS - A quick and dirty usability scale

John Brooke

Redhatch Consulting Ltd.,
12 Beaconsfield Way,
Earley, READING RG6 2UX
United Kingdom

email: john.brooke@redhatch.co.uk

Abstract

Usability does not exist in any absolute sense; it can only be defined with reference to particular contexts. This, in turn, means that there are no absolute measures of usability, since, if the usability of an artefact is defined by the context in which that artefact is used, measures of usability must of necessity be defined by that context too. Despite this, there is a need for broad general measures which can be used to compare usability across a range of contexts. In addition, there is a need for “quick and dirty” methods to allow low cost assessments of usability in industrial systems evaluation. This chapter describes the System Usability Scale (SUS) a reliable, low-cost usability scale that can be used for global assessments of systems usability.

Usability and context

Usability is not a quality that exists in any real or absolute sense. Perhaps it can be best summed up as being a general quality of the **appropriateness to a purpose** of any particular artefact. This notion is neatly summed up by Terry Pratchett in his novel “Moving Pictures”:

“ ‘Well, at least he keeps himself fit,’ said the Archchancellor nastily. ‘Not like the rest of you fellows. I went into the Uncommon Room this morning and it was full of chaps snoring!’

‘That would be the senior masters, Master,’ said the Bursar. ‘I would say they are supremely fit, myself.’

‘Fit? The Dean looks like a man who’s swallowed a bed!’

‘Ah, but Master,’ said the Bursar, smiling indulgently, ‘the word “fit”, as I understand it, means “appropriate to a purpose”, and I would say that the body of the Dean is supremely appropriate to the purpose of sitting around all day and eating big heavy meals.’ The Dean permitted himself a little smile. “ (Pratchett, 1990)

In just the same way, the usability of any tool or system has to be viewed in terms of the context in which it is used, and its appropriateness to that context. With particular reference to information systems, this view of usability is reflected in the current draft international standard ISO 9241-11 and in the European Community ESPRIT project MUSiC (Measuring Usability of Systems in Context) (e.g., Bevan, Kirakowski and Maissel, 1991). In general, it is impossible to specify the usability of a system (i.e., its fitness for purpose) without first defining who are the intended users of the system, the tasks those users will perform with it, and the characteristics of the physical, organisational and social environment in which it will be used.

Since usability is itself a moveable feast, it follows that measures of usability must themselves be dependent on the way in which usability is defined. It is possible to talk of some general classes of usability measure; ISO 9241-11 suggests that measures of usability should cover

- effectiveness (the ability of users to complete tasks using the system, and the quality of the output of those tasks),
- efficiency (the level of resource consumed in performing tasks)
- satisfaction (users' subjective reactions to using the system).

However, the precise measures to be used within each of these classes of metric can vary widely. For example, measures of effectiveness are very obviously determined by the types of task that are carried out with the system; a measure of effectiveness of a word processing system might be the number of letters written, and whether the letters produced are free of spelling mistakes. If the system supports the task of controlling an industrial process producing chemicals, on the other hand, the measures of task completion and quality are obviously going to reflect that process.

A consequence of the context-specificity of usability and measures of usability is that it is very difficult to make comparisons of usability across different systems. Comparing usability of different systems intended for different purposes is a clear case of "comparing apples and oranges" and should be avoided wherever possible. It is also difficult and potentially misleading to generalise design features and experience across systems; for example, just because a particular design feature has proved to be very useful in making one system usable does not necessarily mean that it will do so for another system with a different group of users doing different tasks in other environments.

If there is an area in which it is possible to make more generalised assessments of usability, which could bear cross-system comparison, it is the area of subjective assessments of usability. Subjective measures of usability are usually obtained through the use of questionnaires and attitude scales, and examples exist of general attitude scales which are not specific to any particular system (for example, CUSI (Kirakowski and Corbett, 1988)).

Industrial usability evaluation

The demands of evaluating usability of systems within an industrial context mean that often it is neither cost-effective nor practical to perform a full-blown context analysis and selection of suitable metrics. Often, all that is needed is a general indication of the overall level of usability of a system compared to its competitors or its predecessors. Equally, when selecting metrics, it is often desirable to have measures which do not require vast effort and expense to collect and analyse data.

These sorts of considerations were very important when, while setting up a usability engineering programme for integrated office systems engineering with Digital Equipment Co. Ltd, a need was identified for a subjective usability measure. The measure had to be capable of being administered quickly and simply, but also had to be reliable enough to be used to make comparisons of user performance changes from version to version of a software product.

The need for simplicity and speed came from the evaluation methods being used; users from customer sites would either visit a human factors laboratory, or a travelling laboratory would be set up at the customer site. The users would then work through evaluation exercises lasting between 20 minutes and an hour, at the end of which a subjective measure of system usability would be collected. As can be imagined, after this period of time, users could be very frustrated, especially if they had encountered problems, since no assistance was given. If they were then presented with a long questionnaire, containing in excess of 25 questions it was very likely that they would not complete it and there would be insufficient data to assess subjective reactions to system usability.

SUS - the System Usability Scale

In response to these requirements, a simple usability scale was developed. The System Usability Scale (SUS) is a simple, ten-item scale giving a global view of subjective assessments of usability.

SUS is a *Likert scale*. It is often assumed that a Likert scale is simply one based on forced-choice questions, where a statement is made and the respondent then indicates the degree of agreement or disagreement with the statement on a 5 (or 7) point scale. However, the construction of a Likert scale is somewhat more subtle than this. Whilst Likert scales are presented in this form, the statements with which the respondent indicates agreement and disagreement have to be selected carefully.

The technique used for selecting items for a Likert scale is to identify examples of things which lead to extreme expressions of the attitude being captured. For instance, if one was interested in attitudes to crimes and misdemeanours, one might use serial murder and parking offences as examples of the extreme ends of the spectrum. When these examples have been selected, then a sample of respondents is asked to give ratings to these examples across a wide pool of potential questionnaire items. For instance, respondents might be asked to respond to statements such as "hanging's too good for them", or "I can imagine myself doing something like this".

Given a large pool of such statements, there will generally be some where there is a lot of agreement between respondents. In addition, some of these will be ones where the statements provoke extreme statements of agreement or disagreement among all respondents. It is these latter statements which one tries to identify for inclusion in a Likert scale, since, we would hope that, if we have selected suitable examples, there would be general agreement of extreme attitudes to them. Items where there is ambiguity are not good discriminators of attitudes. For instance, while one hopes that there would be a general, extreme disagreement that "hanging's too good" for those who perpetrate parking offences, there may well be less agreement about applying this statement to serial killers, since opinions differ widely about the ethics and efficacy of capital punishment.

SUS was constructed using this technique. A pool of 50 potential questionnaire items was assembled. Two examples of software systems were then selected (one a linguistic tool aimed at end users, the other a tool for systems programmers) on the basis of general agreement that one was "really easy to use" and one was almost impossible to use, even for highly technically skilled users. 20 people from the office systems engineering group, with occupations ranging from secretary through to systems programmer then rated both systems against all 50 potential questionnaire items on a 5 point scale ranging from "strongly agree" to "strongly disagree".

The items leading to the most extreme responses from the original pool were then selected. There were very close intercorrelations between all of the selected items (± 0.7 to ± 0.9). In addition, items were selected so that the common response to half of them was strong agreement, and to the other half, strong disagreement. This was done in order to prevent response biases caused by respondents not having to think about each statement; by alternating positive and negative items, the respondent has to read each statement and make an effort to think whether they agree or disagree with it.

The System Usability Scale is shown in the next section of this chapter. It can be seen that the selected statements actually cover a variety of aspects of system usability, such as the need for support, training, and complexity, and thus have a high level of face validity for measuring usability of a system.

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree					Strongly agree
1. I think that I would like to use this system frequently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
2. I found the system unnecessarily complex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
3. I thought the system was easy to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
4. I think that I would need the support of a technical person to be able to use this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
5. I found the various functions in this system were well integrated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
6. I thought there was too much inconsistency in this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
7. I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
8. I found the system very cumbersome to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
9. I felt very confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	
10. I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	1	2	3	4	5	

Using SUS

The SU scale is generally used after the respondent has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time.

All items should be checked. If a respondent feels that they cannot respond to a particular item, they should mark the centre point of the scale.

Scoring SUS

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.

To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7, and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

SUS scores have a range of 0 to 100.

The following section gives an example of a scored SU scale.

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree	
1. I think that I would like to use this system frequently	1	2	3	4	5	4
2. I found the system unnecessarily complex	1	2	3	4	5	1
3. I thought the system was easy to use	1	2	3	4	5	1
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5	4
5. I found the various functions in this system were well integrated	1	2	3	4	5	1
6. I thought there was too much inconsistency in this system	1	2	3	4	5	2
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5	1
8. I found the system very cumbersome to use	1	2	3	4	5	1
9. I felt very confident using the system	1	2	3	4	5	4
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5	3

Total score = 22

SUS Score = 22 *22.5 = 55

Conclusion

SUS has proved to be a valuable evaluation tool, being robust and reliable. It correlates well with other subjective measures of usability (eg., the general usability subscale of the SUMI inventory developed in the MUSiC project (Kirakowski, personal communication)). SUS has been made freely available for use in usability assessment, and has been used for a variety of research projects and industrial evaluations; the only prerequisite for its use is that any published report should acknowledge the source of the measure.

Acknowledgements

SUS was developed as part of the usability engineering programme in integrated office systems development at Digital Equipment Co Ltd., Reading, United Kingdom.

References

Bevan, N, Kirakowski, J and Maissel, J, 1991, What is Usability?, in H.-J. Bullinger, (Ed.). *Human Aspects in Computing: Design and use of interactive systems and work with terminals*, Amsterdam: Elsevier.

Kirakowski, J and Corbett, M, 1988, Measuring User Satisfaction, in D M Jones and R Winder (Eds.) *People and Computers IV*. Cambridge: Cambridge University Press.

Pratchett, T., 1990 *Moving Pictures*. London: Gollancz

FORM FOR SESSION DESCRIPTION

1. Identification section

1.1 Evaluation code: _____

1.2 Evaluation code session: _____

1.3 Usability evaluation method employed: _____

1.4 Date: / /

1.5 Time used for evaluation: __h __m. Start time: __h__m End time: __h __m

1.6 Evaluator¹: _____

1.7 User²:

2. List of problems

#	Task ³	Time ⁴	Problem description	Scene ⁵	Scenario ⁶	Impact ⁷	Usability Criteria involved ⁸
1							
2							
3							
4							
5							
...							

¹ Name or code of evaluator conducting the test

² Code of user (never to use names to identify users) being tested. Leave this field empty when using an inspection method.

³ **Task**, or its code used to identify a user task described away, being affected by the usability problem.

⁴ **Time**, (min:sec) used for accomplish (or not) a task.

⁵ Pointers to the data set showing the problem, ex: video session, screen snapshot, URL address for the pages and objects (text, link, image, etc.) when evaluating Web applications.

⁶ **Scenario of use** sequence of user action when the problem occurred

⁷ **Impact**, evaluators rating for the problem. Example of severity scale: critical = 3 /important = 2/ minor = 1.

⁸ Criteria used rating the test. Ex. Nielsen's Heuristic, Ergonomic rule number (according to a given set) to identify the criteria which is concerned by the problem, etc.