Common Industry Format Usability Tests

Nigel Bevan
Serco Usability Services
4 Sandy Lane, Teddington, Middx, TW11 0DU, UK

nbevan@usability.serco.com

ABSTRACT

A Common Industry Format for usability test reports is currently being agreed between major software suppliers and purchasers in an initiative co-ordinated by NIST. The objective is to raise the profile of usability in the procurement process, and to demonstrate the consequent benefits of acquiring products with increased usability. For the reports to be useful, they should contain reliable measures of usability. This requires a carefully designed evaluation procedure and use of appropriate metrics.

BACKGROUND

Usability will only be taken seriously when it is part of the acceptance criteria for a product, and this requires a means to specify usability goals and assess their achievement. In many organisations usability is ignored because there are no objective criteria for usability when developing and procuring products.

A Common Industry Format for usability test reports is currently being agreed between major American software suppliers and purchasers in an initiative supported by NIST (the US National Institute of Standards and Technology) (Blanchard, 1998), and in co-operation with ACM SIGCHI. Although it is not at this stage intended to be the basis for a formal standard, material from relevant standards (such as ISO 9241-11) has been incorporated where appropriate.

In making purchase decisions, companies and organisations have traditionally had little indication of how usable a product would be, how productive its users will be, or how much training and support its users would need. The situation has made it difficult to compare products, to plan for support, or estimate total cost of ownership.

The goals of the NIST initiative are to:

- Encourage software suppliers to work more closely with consumer organisations to understand the needs of the users.

- Develop a common usability reporting format for sharing usability data with consumer organisations.

- Conduct a pilot study to determine how well the usability reporting format works and to determine the value of using this information for software procurement.

The scheme is due to be published and start pilot testing in early 1999. The information in this paper is based on the draft report format which is in the process of being finalised.

AUDIENCE

The CIF is intended to be used by usability professionals within supplier organisations to generate reports that can be used by consumer organisations. The CIF is also meant to be used by consumer organisations to verify that a particular report is CIF compliant. The CIF report is intended for two types of readers:

- Technical professionals and managers who are using the test results to make procurement decisions.

- Human factors or other usability professionals in consumer organisations who are evaluating both the technical merit of usability tests and the usability of the products.

The scope is currently focused on empirical usability testing of commercial application software for end-users. Summative type usability testing is recommended. However the format is intended to support clear and thorough reporting of both the methods and the results of any user-based test.

Who Performs the Test

The usability test may be performed by the software supplier using its own usability group or by contracting with an independent testing facility or a consumer organisation. The consumer organisation interested in purchasing this software may accept the test results or may replicate the test.

In many cases, the supplier organisation will provide the results of the last usability test conducted in the course of software development to the consumer organisation. This allows organisations that have a usability testing program in place to participate without incurring the expense of additional usability testing.

Pilot study

NIST is planning to co-ordinate a 30-month trial of the CIF. Pairs of supplier and consumer companies will enter into an agreement with NIST. An important component of the trial is to collect business data that will demonstrate the benefits of incorporating usability into procurement decision making. Consumer companies will supply NIST with internal data on overhead, productivity and user satisfaction, and record it in a common format in the NIST database.

Report format

The exact contents of the report format are still being finalised, but it is expected that it will require the types of information shown in Annex 1.

Type of test

One of the issues being discussed is the extent to which the report format should require a particular style of test. Many consumer organisations would like reports to contain usability measures that would enable comparisons to be made between products. However, the primary objective of much usability testing is to identify problems rather than to produce measures. A test that is optimised to provide feedback into design may provide little information of any value to a purchaser. The most effective formative usability tests give rapid iterative feedback using three to five participants (Nielsen, 1993). These tests will frequently be exploratory, with the observer discussing the interaction with the participant and giving hints when necessary. Providing the results of such a test to a potential purchaser indicates that attempts have been made to improve the usability of the product, but does not demonstrate whether the product has adequate usability.

The purchaser needs to know how usable the product will be for particular types of users carrying out defined tasks in a specific environment. This is not just a matter of the ease of use of the interface – it is the extent to which the product can meet the needs of the user as a result of providing appropriate functionality, performance, reliability and ease of use (Bevan, 1997). This is the approach to usability taken in ISO 9241-11, and operationalised in the MUSiC methods (Bevan and Macleod, 1994).

Usability and quality in use

To distinguish this approach to usability measurement from narrower concerns with ease of use, the broad objective of meeting user needs is also known as quality in use (Bevan, 1995). ISO/IEC 14598-1 takes this broad view, explaining that quality in use is the user's view of the quality of a system, and is measured in terms of the result of using the system, rather than properties of the system itself. Quality in use is the combined effect of the system quality characteristics for the end user. A product meets the requirements of a particular user if it enables the user to be effective, productive in use of time and resources, and satisfied, regardless of the specific attributes the product possesses. Quality in use is thus synonymous with the approach to usability taken in ISO 9241-11 and the CIF format. It is a black box approach to specification and evaluation, assessing the extent to which a product or service meets the needs of users, irrespective of the specific attributes of the product.

MEASURING USABILITY

Reliable measures of usability can only be obtained if representative users have carried out realistic tasks in realistic environments. This requires a carefully designed evaluation procedure and use of appropriate metrics.

The MUSiC Performance Measurement Method (Macleod et al 1997), based on ISO 9241-11, is an example of how this can be achieved:

1. **Define the product to be tested**. The first step is to define which version and which components of a product or system are to be evaluated.

2. **Define the Context of Use** The next step is to clarify the intended context of use of the product: what are the intended user groups, what skills and what mental and physical capabilities will the intended users have, what task goals can be achieved with the product, and what physical and social conditions will the product be used in? It is important to understand the details of this context in order to provide a basis for evaluation.

   The context in which the product is to be used should be identified in terms of:

   - *the characteristics of the intended users*. Relevant characteristics of the users can include knowledge, skill, experience, education, training, age and physical capabilities. If necessary, define the characteristics of different types of user, for example with different levels of experience or performing different roles or with different capabilities (e.g. maintainers, installers, the elderly).

   - *the tasks the users are to perform*. The description should include the overall goals of use of the system. The characteristics of tasks that can influence quality in use in typical scenarios should be described, e.g. the frequency and the duration of performance. The description should include the allocation of activities and operational steps between the human and technological resources. Tasks should not be described solely in terms of the functions or features provided by a product or system.

   - *the environment in which the users are to use the product*. The description of the hardware, software and materials can be in terms of a set of products, one or more of which can be the focus of human-centred specification or evaluation, or it can be in terms of a set of attributes or performance characteristics of the hardware, software and other materials. Relevant characteristics of the physical and social environment also need to be described. Aspects which may need to be described include attributes of the wider technical environment (e.g. a local area network), the physical environment (e.g. workplace, furniture), the ambient environment (e.g. temperature, humidity), the legislative environment (e.g. laws, ordinances, directives and standards) and the social and cultural environment (e.g. work practices, organisational structure and attitudes).

   An effective way to identify the necessary information is to use an annotated checklist of user, task and environmental characteristics, such as can be found in the Usability Context Analysis Guide (Thomas and Bevan 1996).

   The output from this activity should be a description of the relevant characteristics of the users, tasks and environment which identifies what aspects have an important impact on the system design.

3. **Specify the usability requirements** Usability is measured in specific contexts of use. Measures of usability have three essential components:

   - Effectiveness: can users complete their tasks correctly and completely?

   - Productivity: are tasks completed in an acceptable length of time?

   - Satisfaction: are users satisfied with the interaction?

   Usability requirements are often set by comparison with alternative means of achieving a task. For instance, a new version of a product should enable users to complete their task more effectively in a shorter time or with less support and with more satisfaction than when using existing the existing version.

4. **Specify the Context of Evaluation** so that the evaluation can be carried out in conditions as close as possible to those in which the product will be used. It is important that:

- Users are representative of the population of users who use the product

- Tasks are representative of the ones for which the system is intended

- Conditions are representative of the normal conditions in which the product is used

5. **Design an evaluation** to meet the specified Context of Evaluation. The evaluation measures the performance and satisfaction of users as they perform set tasks within this context. Satisfaction can be measured with a validated questionnaire such as SUMI (Kirakowski, 1996).

   By controlling the context of evaluation, experience has shown that reliable results can be obtained with a sample of only eight participants.

6. **Perform the user tests and collect data.** When assessing quality in use it is important that the users work unaided, only having access to forms of assistance that would be available under normal conditions of use. As well as measuring effectiveness, efficiency and satisfaction it is usual to document the problems users encounter, and to obtain clarification by discussing the problems with users at the end of the session. It is often useful to record the evaluation on video, which permits more detailed analysis, and production of video clips. It is also easier for users to work undisturbed if they are monitored remotely by video.

7. **Analyse and interpret the data**. The data is used to calculate metrics for effectiveness, efficiency, and satisfaction. Effectiveness is calculated by using a scoring system to assess the business value of incomplete or partially correct task outputs. Efficiency is measured as effectiveness divided by the resources used (normally task time), for example giving a measure of output per minute. SUMI scores can be compared to the industry norms for office software.

8. **Produce a usability report**. The CIF format specifies all the information which should be included in a report.

CONCLUSIONS

The CIF format encourages a broad business-oriented approach to usability, reporting the results of tests where representative users have carried out realistic tasks in realistic environments. For reliable results there should be at least eight participants for each user type, but it is also possible to report the results of formative evaluations using smaller numbers of participants and compromising some aspects of the realism.

Once the format is published, anyone will be free to try using it. Purchasers will be encouraged to ask suppliers to provide CIF format usability reports, thus providing a greater incentive for suppliers to provide usable products.

NIST will be collecting data from consumer organisations using the CIF format, to provide evidence of the benefits of improved usability. Trying to relate usability to improved productivity and reduced overheads is an important part of the strategy. Pairs of supplier and purchaser organisations planning to use the format are encouraged to contact NIST (iuwinfo@nist.gov).

As an active UK contributor to the development of the CIF format, I look forward to the CIF being adopted in UK and Europe.

ACKNOWLEDGEMENTS

REFERENCES

Bevan N (1995) Measuring usability as quality of use. Journal of Software Quality, 4, 115-130.

Bevan N (1997) Quality and usability: A new framework. In: van Veenendaal, E, and McMullan, J (eds) Achieving software product quality, Tutein Nolthenius, Netherlands.

Bevan N and Macleod M (1994) Usability measurement in context. Behaviour and Information Technology, **13**, 132-145

Blanchard H (1998) The application of usability testing results as procurement criteria for software. SIGCHI Bulletin, July 1998.

ISO/IEC FCD 9126-1 (1998) Software product quality - Part 1: Quality model.

ISO 9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability.

ISO/IEC 14598-1 (1998) Information Technology - Evaluation of Software Products - Part 1 General guide.

Kirakowski J (1996) The software usability measurement inventory: background and usage. In: P Jordan, B Thomas, & B Weerdmeester, Usability Evaluation in Industry. Taylor & Frances, UK.

Macleod M, Bowden R, Bevan N and Curson I. (1997) The MUSiC Performance Measurement Method. Behaviour and Information Technology, 16.

Nielsen J (1993) Usability Engineering. Academic Press.

Thomas C and Bevan N, eds (1995) Usability Context Analysis: A practical guide, Version 4. Serco Usability Services, Teddington, UK

Annex 1. Likely content of Common Industry Format for usability reports

TITLE PAGE

- identify the report as a Common Industry Format (CIF)

- name of the product and the version that was tested

- when the test was conducted

- contact information for questions about the test

EXECUTIVE SUMMARY

- the identity and a description of the product

- the reason for and nature of the test

- a summary of the method of the test including the number of and type of participants and their tasks

- results expressed as mean scores

- if differences between values or products are claimed, the probability that the difference did not occur by chance

PRODUCT DESCRIPTION

- the formal product name and release or version

- a brief description of the environment in which it should be used

- what parts of the product were evaluated

- the user population for which the product is intended

- the type of user work that is supported by the product

TEST OBJECTIVES

- the objectives for the test and any areas of specific interest

PARTICIPANTS

- the total number of participants tested

- segmentation of user groups tested (if more than one user group was tested)

- the key characteristics and capabilities expected of the user groups being evaluated

- how participants were selected and whether they had the essential characteristics and capabilities

CONTEXT OF PRODUCT USE IN THE TEST

- the task scenarios for testing

- any task data given to the participants,

- any completion or performance criteria established for each task

- type of space in which the evaluation was conducted

- computer hardware model, operating system versions, etc

- if the product uses a web browser, the browser type and version

- if the product has a screen based visual interface, the screen size and monitor resolution

- if the product has an audio interface, relevant settings or values for the audio bits, volume, etc

DESIGN OF THE TEST

- the logical design of the test

- brief definitions of any independent variables and control variables

- brief description of the measures for which data were recorded for each set of conditions

- any policies and procedures for training, coaching, assistance, interventions or responding to questions

- summary of the task instructions (put the exact task instructions in an appendix)

USABILITY METRICS

- the percentage of participants who completely and correctly achieved each task goal

- if it is necessary to provide participants with assists, efficiency and effectiveness metrics must be determined for both unassisted and assisted conditions

- the mean time taken to complete each task, together with the range and standard deviation of times across participants

- summary of other usability results