Who Needs Usability Metrics?

Nigel Bevan and Ian Curson

National Physical Laboratory, Usability Services
Teddington, Middlesex, TW11 0LW, UK
Nigel.Bevan@npl.co.uk

ABSTRACT There is a great deal of debate about the relative value of usability evaluation techniques which produce metrics, versus techniques aimed at rapid problem diagnosis and feedback. Are metrics only of value to researchers, or do they provide real business benefits? Individual opinion tends to be polarised, and to generate champions for either concentrating on measuring usability, or on improving the product through identifying usability defects. The panel will discuss the following issues. When, if ever, is it useful to obtain metrics? Are metrics useful in a business context? What metrics are most useful? Are there inherent dangers in presenting metrics to a statistics-naïve audience?

KEYWORDS usability, metrics, measures, evaluation, cost-effectiveness

"...when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."
(Lord Kelvin, addressing the Institute of Civil Engineers in 1883)

How can you tell whether a system is usable unless you have some way of measuring it? Without measurement, usability remains a matter of opinion. Usability metrics would allow usability to be assessed along with other software metrics. Numerical goals for usability could form part of the product specification and be tracked in a quality system.

But do we know how to measure usability? Do we even know what usability is? (Bevan et al 1991, Bevan 1995). Incomplete or incorrect measures of usability give a false sense of confidence.

One well established approach to usability measurement is the MUSiC method (Bevan and Mac'eod 1994) based on ISO 9241-11 (1997). To obtain these measures requires user testing in controlled conditions with 8 or more users. But is this approach an appropriate use of scarce usability resources during development, when we know that iterative evaluation with 3-5 users is more cost effective for improving usability? (Nielsen 1993).

With resource constraints there is a temptation to measure and quote usability metrics based on 3 or 4 users who may have been using a product in conditions which are far removed from the real world. The resulting figures quoted in reports sound convincing, but how many people have the expertise to judge whether they are meaningful?

Given the appalling usability of many systems in everyday usage, wouldn't we be better to concentrate all our efforts in working to remove the maximum number of usability defects before delivery?

Clare-Marie Karat

IBM TJ Watson Research Center, USA

Clare-Marie has worked for the last fifteen years in industry and government as a user interface designer and human factors engineer. She has published widely on the topic of human factors cost-benefit analysis.

Usability metrics are a key component of usability engineering in successful product development. To develop a successful product, it is critical to understand the goals of the product for the organization developing it and for the product's intended customers and target user populations. It has been my practical experience, that when the goals for a product and the associated

metrics are defined, and the product team commits to them, success is more likely to be achieved. Product goals may address a number of areas (e.g., productivity, usability, professionalism of service delivery, security, customer satisfaction, profit). When the business decision is made to fund product development, the expected business impact of the goals is defined within the organization. A subset of the goals needs to address the usability of the system. Usability metrics to measure progress towards the usability goals may be based on quantitative and qualitative indicators, and the expected business value of achieving these goals is employed in making appropriate tradeoffs during product development. All of the product goals must be understood in terms of their relative importance and the time, money, and resource required to achieve them. It is the common language in which the multidisciplinary product team and senior management in an organization can communicate from the time of product concept to its release. As the product nears release, the financial business value of customer and end-user product goals is the common language in which marketing can communicate the benefits of the product to those who will purchase and use it.

The cost-benefit analysis of product goals and the amount of effort necessary to achieve them in development can be based on subjective judgement or on qualitative and quantitative data that can be shared and discussed amongst the team members and decision makers. For example, consider a situation where the business value of a product to customers (and its competitive advantage in the marketplace) is to be a productivity enhancement to customers' business workflows in meeting their clients requests, and a means of improving end user satisfaction with their work environments. How is the product team to achieve these goals during the development process without the definition of usability goals to address these business goals, the use of usability metrics to measure progress towards the goals, and the use of cost-benefit data based on these measures to make tradeoffs in time, money and resource for competing goals? I will present examples from development situations where usability goals and metrics were employed and situations where they were not part of the development process.

Ian Curson

National Physical Laboratory, UK

Ian is an experienced usability consultant with a background in programming and systems design. He

provides regular training courses in the MUSiC Performance Measurement Method

The goal for commercial systems developers is to meet an agreed specification. If usability targets are not included in that specification, then usability is nothing more than an optional extra - a bonus feature to be considered only in the unlikely event that development progresses ahead of schedule. This should be no great surprise: the specification forms part of the contract and dictates the criteria for success. For usability to be considered, it must be explicitly stated in those criteria. Statements like "usability must be considered" are not good enough, because usability can be considered without doing anything about it. So we need clear goals against which we can compare the system.

For me, the issue is not whether we need metrics - of course we do. Rather, it is to ensure that metrics are meaningful to developers and customers alike, and are quick and inexpensive to apply. Metrics must be meaningful to enable customers to specify reasonable targets, and to enable developers to work towards them. They must be quick and easy to apply because the system should be assessed at key stages throughout development, and not just as part of acceptance testing.

Metrics must also accurately reflect the goals of the system in use, so they must be obtained from a controlled study of real world use. This is best achieved by setting realistic goals for representative users in an appropriate environment and assessing the outcome against customer-supplied quality criteria. Such a scientific yet pragmatic approach often quells the scepticism of developers, who may be expecting what is to them an inconsequential and incomprehensible stream of psychobabble. The results are perceived as useful, and the evaluation method is straightforward enough to be applied by developer and customer alike.

Metrics are essential - but only if they are the right metrics.

Jon Mysel

The Hiser Group, Australia

Jon is a Principal Consultant with The Hiser Group. Jon brings with him 10 years of experience in the specification, design and evaluation of computer system user interfaces.

Usability metrics play a vital role in convincing a potential client of the benefits of usability and user centred design methods. Metrics can present a persuasive argument. For example, some work we recently did for an insurance company on the evaluation

of an existing application led to a two thirds reduction in help desk calls. When attempting to secure funding for usability activities, the ability to show benefits which result in real cost savings go a long way in convincing organisations to invest in usability activities.

However, once the design has begun, the issue is not so much *measurement*, but rather *identification*. The focus of our analysis, design and evaluation activities is to identify potential usability issues and develop appropriate recommendations to resolve them. If one user has difficulty using a part of an application, we need to understand why. It is not necessary to wait for a *statistically significant* number of users to have a problem in order to obtain valuable design insights.

It is not whether you are for or against usability metrics, but when they are helpful. Metrics can illustrate real business benefits. Detailed statistics relating to usability testing is rarely valuable. Statistics tied to real business goals is. Measurement can demonstrate when the usability techniques have delivered real value. By tracking the improvements in real numbers such as reduced help desk calls, greater productivity, increased sales. In other instances it is more important to identify issues than it is to measure them. If you are trying to develop successful designs, you need to understand what the issues are and when users are having difficulty. By identifying these issues early in the design process, you can quickly, and effectively resolve them.

Jurek Kirakowski

Human Factors Research Group, Ireland

Jurek is Director of the Human Factors Research Group at University College, Cork, and originator of psychometric questionnaires, including SUMI, for assessing perceived usability.

Many 'metrics' and 'measures' of software quality, complexity, as well as of process and organisational characteristics have been proposed. By and large, most of these metrics and measures fail the two principal requirements that they should be reliable and valid.

A key question in the production of metrics is the kind of activity that is carried out in order to generate them:

Testing with end users This is the only way of finding out how end users will react to the software when it is released, and may reveal problems with the software that cannot be detected by inspections or appeal to the use of standards. But you have to deal

with people in order to derive the metrics, which is often difficult for a technology-orientated company to take on board, and you have to understand a lot about the software in relation to its intended context of use in order to be able to test it in realistic conditions.

Testing with expert assessors These are often consultants, part of whose professional expertise is fitting in with the goals and the work practices of the employing organisation. The extent of an expert's knowledge base usually enables them to summarise comparative judgements about the software in metric-like data which is easily assimilated by management. But expert assessors are usually expensive, often committed to a particular approach or philosophy, and usually offer no index of how reliable their quantitative assessments are.

Testing against a theoretical model This kind of testing is usually supplied as a software suite which fits in with the testing practices of most organisations (ie validating against known good and bad input cases) and therefore fits in well with the work practices of most organisations developing software. But by definition such approaches are theory-driven, most tools to support such evaluations are time-consuming, and the output metrics may be difficult to relate to actual user experience with the software.

The biggest problem with most proposed metrics is that there are no standards regarding what one should expect a 'good' software product to score on the metric: no real baseline data.

Metrics emerging from the 'softest' kinds of activities (testing with end users) may hold the brightest ray of hope for the future. Such metrics can be validated and have their reliability assessed using psychometric models established by psychologists in a work effort that spans most of the 20th century.

REFERENCES

- Bevan N (1995) Usability is quality of use. In: Anzai & Ogawa (eds) Proc. 6th International Conf. on Human Computer Interaction, July 1995. Elsevier.
- Bevan N, Kirakowski J and Maissel J (1991) What is usability? In: Bullinger HJ (1991) Proc. 4th International Conf. on HCI, Stuttgart. Elsevier.
- Bevan N, Macleod M (1994) Usability measurement in context. Behaviour and Information Technology. 13, 132-145.
- ISO (1997) ISO DIS 9241-11 Guidance on Usability.
- Nielsen J (1993) Usability Engineering. Academic Press.