

Classifying and selecting UX and usability measures

Nigel Bevan

Professional Usability Services

12 King Edwards Gardens, London W3 9RG, UK

mail@nigelbevan.com

www.nigelbevan.com

ABSTRACT

There are many different types of measures of usability and user experience (UX). The overall goal of usability from a user perspective is to obtain acceptable effectiveness, efficiency and satisfaction (Bevan, 1999, ISO 9241-11). This paper summarises the purposes of measurement (summative or formative), and the measures of usability that can be taken at the user interface level and at the system level. The paper suggests that the concept of usability at the system level can be broadened to include learnability, accessibility and safety, which contribute to the overall user experience. UX can be measured as the user's satisfaction with achieving pragmatic and hedonic goals, and pleasure.

WHY MEASURE UX/USABILITY?

The most common reasons for measuring usability in product development are to obtain a more complete understanding of users' needs and to improve the product in order to provide a better user experience.

But it is also important to establish criteria for UX/usability goals at an early stage of design, and to use summative measures to evaluate whether these have been achieved during development.

Summative measures

Summative evaluation can be used to establish a baseline, make comparisons between products, or to assess whether usability requirements have been achieved. For this purpose, the measures need to be sufficiently valid and reliable to enable meaningful conclusions to be drawn from the comparisons. One prerequisite is that the measures are taken from an adequate sample of typical users carrying out representative tasks in a realistic context of use. Any comparative figures should be accompanied by a statistical assessment of whether the results may have been obtained by chance.

For example, the test method for everyday products in ISO 20282-2 points out that to obtain 95% confidence that 80% of users could successfully complete a task would for example require 28 out of 30 users tested to be successful. If 4 out of 5 users in a usability test were successful, even if the testing protocol was perfect there is

20% chance that the success rate for a large sample of users might only be 51%.

Although summative measures are most commonly obtained from user performance and satisfaction, summative data can also be obtained from hedonic questionnaires (e.g. Hassenzahl et al., 2003; Lavie and Tractinsky, 2004) or from expert evaluation, such as the degree of conformance with usability guidelines (see for example Jokela, et al, 2006).

Formative measures

Formative evaluation can be used to identify UX/usability problems, to obtain a better understanding of user needs and to refine requirements. The main data from formative evaluation is qualitative. When formative evaluation is carried out relatively informally with small numbers of users, it does not generate reliable data from user performance and satisfaction.

However some measures of the product obtained by formative evaluation, either with users or by an expert, such as the number of problems identified, may be useful, although they should be subject to statistical assessment if they are to be interpreted.

In practice, even when the main purpose of an evaluation is summative, it is usual to collect formative information to provide design feedback at the same time.

WHAT MEASURES SHOULD BE USED?

There are two types of UX/usability measures: those that measure the result of using the whole system (usability in use) and measures of the quality of the user interface (interface usability).

SYSTEM USABILITY

ISO 9241-11 (1998) defines usability as:

the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use

and ISO 9241-171 (2008) defines accessibility as:

usability of a product, service, environment or facility by people with the widest range of capabilities

These definitions mean that for a product to be usable and accessible users should be able to use a product or web site to achieve their goals in an acceptable amount of time, and be satisfied with the results. ISO/IEC standards for software quality refer to this broad view of usability as “quality in use”, as it is the user’s overall experience of the quality of the product (Bevan, 1999). This is a black-box view of usability: what is achieved, rather than how.

The new draft ISO standard ISO/IEC CD 25010.2 (2008) proposes a more comprehensive breakdown of quality in use into *usability in use* (which corresponds to the ISO 9241-11 definition of usability as effectiveness, efficiency and satisfaction); *flexibility in use* (which is a measure of the extent to which the product is usable in all potential contexts of use, including accessibility); and *safety* (which is concerned with minimising undesirable consequences):

Quality in use

Usability in use

- Effectiveness in use
- Productivity in use
- Satisfaction in use
 - Likability (satisfaction with pragmatic goals)
 - Pleasure (satisfaction with hedonic goals)
 - Comfort (physical satisfaction)
 - Trust (satisfaction with security)

Flexibility in use

- Context conformity in use
- Context extendibility in use
- Accessibility in use

Safety

- Operator health and safety
- Public health and safety
- Environmental harm in use
- Commercial damage in use

Usability in use is similar to the ISO 9241-11 definition of usability:

- *Effectiveness*: “accuracy and completeness.” Error-free completion of tasks is important in both business and consumer applications.
- *Efficiency*: “resources expended.” How quickly a user can perform work is critical for business productivity.
- *Satisfaction*: the extent to which expectations are met. Satisfaction is a success factor for any products with discretionary use; it’s essential for maintaining workforce motivation.

Usability in use also explicitly identifies the need for a product to be usable in the specified contexts of use:

- *Context conformity*: the extent to which usability in use meets requirements in all the required contexts of use.

Flexibility in use: the extent to which the product is usable in all potential contexts of use:

- *Context conformity in use*: the degree to which usability in use meets requirements in all the intended contexts of use.
- *Context extendibility in use*: the degree of usability in use in contexts beyond those initially intended.
- *Accessibility in use*: the degree of usability in use for users with specified disabilities.

Safety: acceptable levels of risk of harm to people, business, data, software, property or the environment in the intended contexts of use.

Safety is concerned with the potential adverse consequences of not meeting the goals. For instance in Cockton’s (2008) example of designing a van hire system, from a business perspective, what are the potential consequences of:

- Not offering exactly the type of van preferred by a potential user group?
- The user mistakenly making a booking for the wrong dates or wrong type of vehicle?
- The booking process taking longer than with competitor systems?

For a consumer product or game, what are the potential adverse consequences of a lack of pleasurable emotional reactions or of achievement of other hedonic goals?

SYSTEM USABILITY MEASURES

Usability in use and flexibility in use are measured by effectiveness (task goal completion), efficiency (resources used) and satisfaction. The relative importance of these measures depends on the purpose for which the product is being used (for example in some personal situations, resources may not be important).

Table 1 illustrates how the measures of effectiveness, resources, safety and satisfaction can be selected to measure quality in use from the perspective of different stakeholders.

From an organisational perspective, quality in use and usability in use is about achievement of task goals. But for the end user there are not only pragmatic task-related “do” goals, but also hedonic “be” goals (Carver & Scheier, 1998). For the end user, effectiveness and efficiency are the do goals, and stimulation, identification, evocation and pleasure are the be goals.

Additional derived user performance measures (Bevan, 2006) include:

- *Partial goal achievement*. In some cases goals may be only partially achieved, producing useful but suboptimal results.

- *Relative user efficiency.* How long a user takes in comparison with an expert.
- *Productivity.* Completion rate divided by task time, which gives a classical measure of productivity.

Table 1. Stakeholder perspectives of quality in use

Stakeholder:	End User <i>Usability</i>	Usage Organisation <i>Cost-effectiveness</i>	Technical support <i>Maintenance</i>
Goal: Characteristic	Personal goals	Task goals	Support goals
System effectiveness	User effectiveness	Task effectiveness	Support effectiveness
System resources	Productivity (time)	Cost efficiency (money)	Support cost
Safety	Risk to user (health and safety)	Commercial risk	System failure or corruption
Stakeholder satisfaction	Hedonic and pragmatic satisfaction	Management satisfaction	Support satisfaction

User satisfaction measures

User satisfaction can be measured by the extent to which users have achieved their pragmatic and hedonic goals. ISO/IEC CD 25010.2 suggests the following types of measure:

- *Likability:* the extent to which the user is satisfied with their perceived achievement of pragmatic goals, including acceptable perceived results of use and consequences of use.
- *Pleasure:* the extent to which the user is satisfied with their perceived achievement of hedonic goals of stimulation, identification and evocation (Hassenzahl, 2003) and associated emotional responses (Norman’s (2004) visceral category).
- *Comfort:* the extent to which the user is satisfied with physical comfort.
- *Trust:* the extent to which the user is satisfied that the product will behave as intended.

Satisfaction is most often measured using a questionnaire. Psychometrically designed questionnaires will give more reliable results than ad hoc questionnaires (Hornbaek, 2006).

Safety and risk measures

There are no simple measures of safety. Historical measures can be obtained for the frequency of health and safety, environmental harm and security failures. A

product can be tested in situations that might be expected to increase risks. Or risks can be estimated in advance.

Evaluation of data from usage of an existing system

Measures of effectiveness, efficiency and satisfaction can also be obtained from usage of an existing system.

Web Metrics

Web-based logs contain potentially useful data that can be used to evaluate usability by providing data such as entrance and exit pages, frequency of particular paths through the site, and the extent to which search is successful. (Burton and Walther, 2001), although it is very difficult to track individual user behaviour (Groves, 2007) without some form of pagetagging combined with pop-up questions when the system is being used, so that the results can be related to particular user groups and tasks.

Application Instrumentation

Data points can be built into code that "count" when an event occurs (for example in Microsoft Office (Harris, 2005)). This could be the frequency with which commands are used or the number of times a sequence results in a particular type of error. The data is sent anonymously to the development organization. This real-world data from large populations can help guide future design decisions.

Satisfaction Surveys

Satisfaction questionnaires distributed to a sample of existing users provide an economical way of obtaining feedback on the usability of an existing product or system.

USER INTERFACE USABILITY

The broad quality in use perspective contrasts with the narrower interpretation of usability as the attributes of the user interface that makes the product easy to use. This is consistent with one of the views of usability in HCI, for example in Nielsen’s (1993) breakdown where a product can be usable, even if it has no utility (Figure 1).

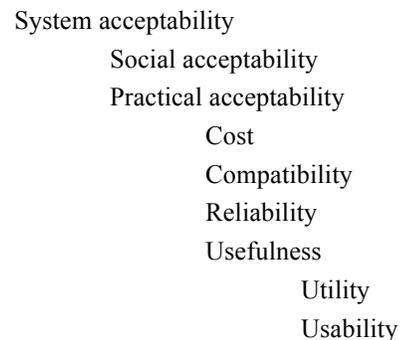


Figure 1. Nielsen’s categorisation of usability

User interface usability is a pre-requisite for system usability.

Expert-based methods

Expert evaluation relies on the expertise of the evaluator, and may involve walking through user tasks or assessing conformance to UX/usability guidelines or heuristics.

Measures that can be obtained from expert evaluation include:

- Number of violations of guidelines or heuristics.
- Number of problems identified.
- Percentage of interface elements conforming to a particular guideline.
- Whether the interface conforms to detailed requirements (for example the number of clicks required to achieve specific goals).

If the measures are sufficiently reliable, they can be used to track usability during development.

Automated evaluation methods

There are some automated tools (such as WebSAT and LIFT) that automatically test for conformance with basic usability and accessibility rules. Although the measures obtained are useful for screening for basic problems, they only test a very limited scope of usability issues (Ivory & Hearst, 2001).

MEASURING UX, USABILITY AND ACCESSIBILITY

Usability is variously interpreted as good user interface design (ISO 9126-1), an easy to use product (e.g. Cockton, 2004), good user performance (e.g. Väänänen-Vainio-Mattila et al, 2008), good user performance and

satisfaction (e.g. ISO 9241-11), or good user performance and user experience (e.g. ISO 9241-210).

Accessibility may refer to product capabilities (“technical accessibility”) or a product usable by people with disabilities (e.g. ISO 9241-171).

UX has even more interpretations. ISO CD 9241-210 defines user experience as:

all aspects of the user’s experience when interacting with the product, service, environment or facility.

This definition can be related to different interpretations of UX:

- **UX attributes** such as aesthetics, designed into the product to create a good user experience.
- The user’s pragmatic and hedonic **UX goals** (individual criteria for user experience) (Hassenzahl, 2003).
- The actual **user experience** when using the product (this is difficult to measure directly).
- The measurable **UX consequences** of using the product: pleasure, and satisfaction with achieving pragmatic and hedonic goals.

Table 2 shows how measures of system usability and UX are dependent on product attributes that support different aspects of user experience. In Table 2 the columns are the quality characteristics that contribute to the overall user experience, with the associated product attributes needed to achieve these qualities.

Table 2. Factors contributing to system usability and UX

Quality characteristic	UX	Functionality	User interface usability	Learnability	Accessibility	Safety
Product attributes	Aesthetic attributes	Appropriate functions	Good UI design (easy to use)	Learnability attributes	Technical accessibility	Safe and secure design
UX pragmatic do goals	To be effective and efficient					
UX hedonic be goals	Stimulation, identification and evocation					
UX: actual experience	Visceral	Experience of interaction				
Usability (= performance in use measures)	Effectiveness and Productivity in use: effective task completion and efficient use of time			Learnability in use: effective and efficient to learn	Accessibility in use: effective and efficient with disabilities	Safety in use: occurrence of unintended consequences
Measures of UX consequences	Satisfaction in use: satisfaction with achieving pragmatic and hedonic goals					
	Pleasure	Likability and Comfort				Trust

The users' goals may be pragmatic (to be effective and efficient), and/or hedonic (stimulation, identification and/or evocation).

Although UX is primarily about the actual experience of usage, this is difficult to measure directly. The measurable consequences are the user's performance, satisfaction with achieving pragmatic and hedonic goals, and pleasure.

User performance and satisfaction is determined by qualities including attractiveness, functionality and interface usability. Other quality characteristics will also be relevant in determining whether the product is learnable, accessible, and safe in use.

Pleasure will be obtained from both achieving goals, and as a direct visceral reaction to attractive appearance (Norman, 2004).

WHAT SHOULD BE MEASURED?

In a systems development environment, UX/usability measures need to be prioritised:

1. At a high level, whose stakeholder goals are the main concern (e.g. users, staff or managers)?
2. What aspects of effectiveness, efficiency, satisfaction, flexibility, accessibility and safety are most important for these stakeholders?
3. What are the risks if the goals for effectiveness, efficiency, satisfaction, flexibility, accessibility and safety are not achieved in the intended contexts of use?
4. Which of these UX/system usability measures are important enough to validate using user-based testing and/or questionnaires, and how should the users, tasks and measures be selected?
5. Are baseline measures needed to establish requirements? (Whiteside et al, 1998)
6. Which aspects of interface usability can be measured during development by expert evaluation to help develop a product that achieves the UX/system usability goals for the important stakeholders in the important contexts of use?
7. How can UX/usability be monitored during use?

CONCLUSIONS

Discussion of UX and selection of appropriate UX measures would be simplified if the different perspectives on UX were identified and distinguished. The current interpretations of "UX" are even more diverse than those of "usability".

This paper proposes a common framework for classifying usability and UX measures, showing how they relate to broader issues of effectiveness, efficiency, satisfaction, accessibility and safety. It is anticipated that the framework

could be elaborated to incorporate new conceptual distinctions as they emerge.

Understanding how different aspects of user experience relate to usability, accessibility, and broader conceptions of quality in use, will help in the selection of appropriate measures.

REFERENCES

- [1] Bevan, N. (1999) Quality in use: meeting user needs for quality, *Journal of Systems and Software*, 49(1), pp 89-96.
- [2] Bevan, N. (2006) Practical issues in usability measurement. *Interactions* 13(6): 42-43
- [3] Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.
- [4] Burton, M and Walther, J (2001) The value of web log data in use-based design and testing. *Journal of Computer-Mediated Communication*, 6(3). jcmc.indiana.edu/vol6/issue3/burton.html
- [5] Cockton, G. (2004) From Quality in Use to Value in the World. *CHI 2004*, April 24–29, 2004, Vienna, Austria.
- [6] Cockton, G. (2008a) Putting Value into E-valuation. In: *Maturing Usability. Quality in Software, Interaction and Value*. Law, E. L., Hvannberg, E. T., Cockton, G. (eds). Springer.
- [7] Cockton (2008b) What Worth Measuring is. *Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, Reykjavik, Iceland.
- [8] Groves, K (2007). The limitations of server log files for usability analysis. *Boxes and Arrows*. www.bboxesandarrows.com/view/the-limitations-of
- [9] Harris, J. (2005) An Office User Interface Blog. <http://blogs.msdn.com/jensenh/archive/2005/10/31/487247.aspx> Retrieved January 2008.
- [10] Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, 13, 479-497.
- [11] Hassenzahl, M. (2003) The thing and I: understanding the relationship between user and product. In *Funology: From Usability to Enjoyment*, M. Blythe, C. Overbeeke, A.F. Monk and P.C. Wright (Eds), pp. 31 – 42 (Dordrecht: Kluwer).
- [12] Hornbaek, K (2006). Current practices in measuring usability. *Int. J. Human-Computer Studies* 64 (2006) 79–102

- [13] ISO 9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on Usability. ISO.
- [14] ISO FDIS 9241-171 (2008) Ergonomics of human-system interaction -- Part 171: Guidance on software accessibility. ISO.
- [15] ISO CD 9241-210 (2008) Ergonomics of human-system interaction -- Part 210: Human-centred design process for interactive systems. ISO.
- [16] ISO 13407 (1999) Human-centred design processes for interactive systems. ISO.
- [17] ISO TS 20282-2 Ease of operation of everyday products -- Part 2: Test method for walk-up-and-use products. ISO.
- [18] ISO/IEC 9126-1 (2001) Software engineering - Product quality - Part 1: Quality model. ISO.
- [19] ISO/IEC CD 25010.2 (2008) Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Quality model
- [20] Ivory, M.Y., Hearst, M.A. (2001) State of the Art in Automating Usability Evaluation of User Interfaces. ACM Computing Surveys, 33,4 (December 2001) 1-47. Accessible at <http://webtango.berkeley.edu/papers/ue-survey/ue-survey.pdf>
- [21] Jokela, T., Koivumaa, J., Pirkola, J., Salminen, P., Kantola, N. (2006) “Methods for quantitative usability requirements: a case study on the development of the user interface of a mobile phone”, *Personal and Ubiquitous Computing*, 10, 345 – 355.
- [22] Nielsen, J. (1993) *Usability Engineering*. Academic Press.
- [22] Norman, D. (2004) *Emotional design: Why we love (or hate) everyday things* (New York: Basic Books).
- [23] Väänänen-Vainio-Mattila, K., Roto, V., Hassenzahl, M. (2008) *Towards Practical UX Evaluation Methods. Proceedings of Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, Reykjavik, Iceland.
- [24] Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). *Usability engineering: Our experience and evolution*. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (1st Ed.) (pp. 791–817). North-Holland.