

The evaluation of software usability

Andrew Dillon

This item is not the definitive copy. Please use the following citation when referencing this material: Dillon, A. (2001) The evaluation of software usability. In: W. Karwowski (ed). Encyclopedia of Human Factors and Ergonomics. London: Taylor and Francis.

1. Introduction

Usability is a measure of interface quality that refers to the effectiveness, efficiency and satisfaction with which users can perform tasks with a tool. Evaluating usability is now considered an essential part of the system development process and a variety of methods and have been developed to support the human factors professional in this work.

2. The concept of usability

Historically, the concept of usability has been defined in multiple ways, usually on one of the following bases:

- **Semantics:** in this case usability is equated to terms such as ‘ease of use’ or ‘user-friendliness’, without formal definition of the properties of the construct.
- **Features:** here, usability is equated to the presence or absence of certain features in the user interface such as Windows, Icons, Menus or Pointing devices
- **Operations:** where the term is defined in terms of performance and affective levels manifest by users for certain task and environmental scenarios.

The first type of definition was common in the 1970s but has proved largely useless for design purposes since it offers neither useful guidance for designers nor perspective for evaluators. The feature-based approach stems from the longstanding desire to specify in advance desirable interface attributes that enhance usability. However, this type of definition rests on an assumption that usability is an inherent part of the application. This assumption is false since one could always envisage a combination of users, with certain task demands, in a particular environment, for whom a given set of features would be sub-optimal. Recognizing this, most human factors professionals now employ an operational definition in their work.

Shackel (1991) is the major developer of this operational approach. He defined usability as the artifact's capability, in human functional terms, to be used easily, effectively and satisfactorily by specific users, performing specific tasks, in specific environments. The essence of the operational definition is that it explicitly places usability at the level of the interaction between users and the artifact. This takes it beyond the typical features-based

definitions common in the field. Furthermore, in setting criteria for assessing usability, this approach better supports the evaluation of any tool and the subsequent interpretation of the test results. Usability therefore refers not to a set of interface features, but to a context-dependent measure of human-computer *interaction*.

3. Evaluating usability

There exist multiple methods of evaluating usability depending on available resources (time facilities and labor), evaluator experience, ability and preference, and the stage of development of the tool under review. In broad terms it is worth making the following distinctions between evaluation methods:

1. User-based: where a sample of the intended users try to use the application
2. Expert-based: where an HCI or usability expert makes an assessment of the application
3. Model-based: where an HCI expert employs formal methods to predict one or more criteria of user performance

3.1 User-based methods

Testing an application with a sample of users performing a set of pre-determined tasks is generally considered to yield the most reliable and valid estimate of an application's usability. Performed either in a usability test laboratory or a field site, the aim of such a test is to examine the extent to which the application supports the intended users in their work. Tightly coupled to the operational approach to usability definition, the user-based approach draws heavily on the experimental design tradition of human factors psychology in employing task analysis, pre-determined dependent variables and, usually, quantitative analysis of performance supplemented with qualitative methods.

In a typical user-based evaluation, test subjects are asked to perform a set of tasks with the technology. Depending on the primary focus of the evaluator, the users' success at completing the tasks and their speed of performance may be recorded. After the tasks are completed, users are often asked to provide data on likes and dislikes through a survey or interview, or may be asked to view with the evaluator part of their own performance on video and to describe in more detail their performance and perceptions of the application. In this way, measures of effectiveness, efficiency and satisfaction can be derived, problems can be identified and re-design advice can be determined. In certain situations, concurrent verbal protocols might be solicited to shed light on users' thought processes while interacting with the tool so that issues of comprehension and user cognition can be addressed. In a usability lab, the complete interaction is normally video recorded for subsequent analysis of transactions, navigation, problem handling etc. However more informal approaches are also possible. Some user-based tests are unstructured, involving the user and the evaluator jointly interacting with the system to gain agreement on what works and what is problematic with the design. Such participative approaches can be very

useful for exploring interface options in the early stages of design where formal quantitative assessments might be premature.

In an ideal world user testing with a large sample of the intended user population would routinely occur, however due to resource limitations, user-based tests are often constrained. As a result, there is considerable interest among HCI professionals in determining how to gain the most information from the smallest sample of users. While popular myths exist about being able to determine a majority of problems with only 2 or 3 users, Lewis (1994) has shown that the sample size requirement is largely dependent on the type of errors one seeks to identify and their relative probability of occurrence. Whereas 3 users might identify many problems in a new application, substantially more users will be required to tease out the remaining problems in a mature or revised product.

3.2 Expert-based methods

Expert-based methods refers to any form of usability evaluation which involves an HCI expert examining the application and estimating its likely usability for a given user population. In such cases, users are not employed and the basis for the evaluation lies in the interpretation and judgement of the evaluator. There is considerable interest in this form of evaluation since it can produce results faster and presumably cheaper than user-based tests.

In HCI, two common expert-based usability evaluation methods are Heuristic evaluation (e.g., Nielsen, 1994), and Cognitive Walkthrough (Wharton et al, 1994). Both methods aim to provide evaluators with a structured method for examining and reporting problems with an interface. The Heuristic method provides a simple list of design guidelines which the evaluator uses to examine the interface screen by screen and while following a typical path through a given task. The evaluator reports violations of the guidelines as likely user problems. In the Cognitive Walkthrough method, the evaluator first determines the exact sequence of correct task performance, and then estimates, on a screen by screen basis, the likely success or failure of the user in performing such a sequence. In both methods, the expert must make an informed guess of the likely reaction of users and explain why certain interface attributes are likely to cause users difficulties.

These methods differ in their precise focus. Heuristic methods are based on design guidelines and ultimately reflect the expert's judgement of how well the interface conforms to good design practice. The Cognitive Walkthrough method concentrates more on the difficulties users may experience in learning to operate an application to perform a given task. In practice, usability evaluators tend to adapt and modify such methods to suit their purpose and many experts who perform such evaluations employ a hybrid form of the published methods.

3.3 Model-based methods

Model-based approaches to usability evaluation are the least common form of evaluation but several methods have been proposed which can accurately predict certain aspects of

user performance with an interface such as time to task completion or difficulty of learning a task sequence. In such cases, the evaluator determines the exact sequence of behaviors a user will exhibit through detailed task analysis, applies an analytical model to this sequence and calculates the index of usability.

The most common model-based approach to estimating usability is the GOMS method of Card et al (1983), a cognitive psychology-derived framework that casts user behavior into a sequence of fundamental units (such as moving a cursor to given screen location or typing a well-practiced key sequence) which are allocated time estimates for completion based on experimental findings of human performance from psychology. In this way, any interface design can be analyzed to give an estimate of expert users' time to complete a task. The model has shown itself to be robust over repeated applications (see e.g., Gray et al 1993), though it is limited to predicting time and only then, for error-free performance in tasks involving little or no decision making.

4. Comparisons of methods

The relative advantages and disadvantages of each broad method are summarized in Table 1. Since usability evaluators are trying to estimate the extent to which real users can employ an application effectively, efficiently and satisfactorily, properly executed user-based methods are always going to give the truest estimate. However, the usability evaluator does not always have the necessary resources to perform such evaluations and therefore other methods must be used.

Usability method	ADVANTAGES	DISADVANTAGES
User-based	Most realistic estimate of usability Can give clear record of important problems	Time consuming Costly for large sample of users Requires prototype to occur
Expert-based	Cheap Fast	Expert-variability unduly affects outcome May overestimate true number of problems
Model-based	Provides rigorous estimate of usability criterion Can be performed on interface specification	Measures only one component of usability Limited task applicability

Table 1 - Relative advantages and disadvantages of each usability evaluation method

Comparisons of evaluation methods have recently been reported by HCI professionals but few firm conclusions can yet be drawn. John and Marks (1997) compared multiple evaluation methods and concluded that no one method is best and all evaluation methods

are of limited value. Andre et al (1999) attempted a meta-analysis of 17 comparative studies and remarked that a robust meta-analysis was impossible due to the failure of many evaluation comparisons to provide sufficient statistics. Caveats noted, several practical findings follow:

- It is generally recognized that expert based evaluations employing the heuristic method locate more problems than other methods, including user-based tests. This may suggest that heuristic approaches label as problems many interface attributes that users do not experience as problems or are able to work around.
- The skill-level of the evaluator performing the expert-based method is important. Nielsen (1993) reports that novice evaluators identify significantly fewer problems than experienced evaluators, and both of these groups identify fewer than evaluators who are both expert in usability testing and the task domain for which the tool under review is being designed.
- Team or multiple expert evaluations produce better results than single expert evaluations.

Finally, there are good reasons for thinking that the best approach to evaluating usability is to combine methods e.g., using the expert-based approach to identify problems and inform the design of a user-based test scenario, since the overlap between the outputs of these methods is only partial, and a user-based test normally cannot cover as much of the interface as an expert-based method. Obviously, where usability evaluation occurs throughout the design process, the deployment of various methods at different stages is both useful and likely to lead to greater usability in the final product.

5. Conclusion

Usability evaluation is a core component of user-centered systems design and an essential competency for Human Factors professionals working in the software domain. Test methods vary from laboratory studies of user performance to model-based predictions based on an examination of the interface specification. Choosing among methods is largely a matter of determining what information is needed and at what stage in the development cycle is the evaluation to occur. It will be difficult for an expert examining a prototype to predict user satisfaction, for example, or for a user to reliably estimate her own efficiency from an interface specification, hence the need for more than one type of evaluation method. Clearly, the ultimate test is the behavior of real users interacting under normal working conditions and any single usability evaluation method is an attempt to predict some or all of the issues that will occur in real use. Improvements in evaluation methodology are therefore tied directly to increases in the theoretical analysis of the determinants of use.

References

- Andre, T., Williges, R. and Hartson, H. (1999) The effectiveness of usability evaluation methods: determining the appropriate criteria. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting (Santa Monica CA: HFES) 1090-1094
- Card, S., Moran, T. and Newell, A. (1983) The Psychology of Human-Computer Interaction. (Hillsdale NJ: LEA.)
- Gray, W. John, B. and Atwood, M.(1992) The precis of project Ernestine or An overview of a validation of GOMS. Proc. of the Annual ACM SIGCHI Conference:CHI'92, New York: ACM Press 307-312.
- John B. and Marks, S. (1997) Tracking the effectiveness of usability evaluation methods, Behaviour and Information Technology, 16, 4/5, 188-202.
- Lewis, J (1994) Sample sizes for usability studies: additional considerations. Human Factors, 36(2) 368-378.
- Nielsen, J. (1992) Finding usability problems through heuristic evaluation Proceedings of the ACM SIGCHI Conference: CHI '92 (New York: ACM), 373-380.
- Nielsen, J. (1994) Heuristic Evaluation. In In J. Nielsen and R. Mack (eds.) Usability Inspection Methods (New York: Wiley) 25-62.
- Shackel, B. (1991) Usability - Context, Framework, Definition, Design and Evaluation. In B. Shackel and S. Richardson (eds.) Human Factors for Informatics Usability (Cambridge: Cambridge University Press) 21-38.
- Wharton, C. Rieman, J., Lewis, C. and Polson, P. (1994) The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R. Mack (eds.) Usability Inspection Methods (New York: Wiley) 105-140.

Author:

Andrew Dillon, HCI Lab, SLIS, Indiana University, Bloomington IN 47405, ph: (812 855 3249, email: adillon@indiana.edu)