1998-01-01

Strategies for Evaluating Software Usability

Ronan Fitzpatrick
Dublin Institute of Technology, ronan.fitzpatrick@comp.dit.ie

Recommended Citation
Fitzpatrick, R.: Strategies for Evaluating Software Usability. Methods, Volume: 353, Issue: 1. 1998.

Strategies for Evaluating Software Usability

Ronan Fitzpatrick
Department of Mathematics, Statistics and Computer Science,
Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland.
Tel: +353 (1) 4024610, Fax: +353 (1) 4024994
Email: rfitzpatrick@maths.kst.dit.ie

ABSTRACT
This paper presents a usage analysis and taxonomy of methods which are used to evaluate the usability of computer systems. To accommodate the analysis and taxonomy, a matrix of strategies which can be used for effective usability evaluation is presented. Such an analysis, taxonomy and strategies support human-computer interaction (HCI) professionals who have the responsibility for ensuring computer system usability. The strategies outlined are named Virtual Engineering, Soft Modelling, Hard Review and Real World. This paper also uses a composite set of existing popular generic evaluation methods which can be used as part of these strategies. The methods used are observation, questionnaire, interview, empirical methods, user groups, cognitive walkthroughs, heuristic methods, review methods and model methods. The paper continues by presenting a Usage Analysis Table of these methods and concludes by grouping them into a Taxonomy of Usability Evaluation Methods. A key emphasis of this paper is the appropriateness of individual methods to lifecycle timing.

Keywords
Human-Computer Interaction, (HCI), strategies, usability, usability evaluation, usability evaluation method, lifecycle.

INTRODUCTION
The usability evaluation of computer systems is an important aspect of system development and acquisition which is of interest to students, IS professional and strategic managers. There are a number of usability evaluation methods available to HCI professionals and each of these methods has different characteristics. The choice of appropriate methods and deciding at what stage of the lifecycle they should be applied in order to achieve the highest quality system, can be confusing. Therefore, a classification and analysis of usability methods is an essential usability evaluation tool. However, the classification of usability evaluation methods has presented some difficulty in the past (Dix *et al*. 1998; Reiterer and Oppermann 1993; Whitefield *et al*. 1991). This difficulty has been further compounded by the absence of a universally accepted set of names for generic evaluation methods. This paper addresses these issues by identifying the domain of generic methods that exist, clarifying their classification and presenting an analysis of these methods. Acknowledged confusion and difficulty regarding classification is resolved by restructuring two classifications from respected sources. These restructured classifications add additional perspective to our understanding of usability evaluation. In particular they emphasis the relevance of lifecycle timing. They also contain enhancements by incorporating latest International Organisation for Standardisation guidelines for usability evaluation. They particularly address issues of interest to strategic managers and IS professionals and make reference to current research and practice especially to the ever critical factor of cost.

In this paper the expression "generic method" is used to describe an individual method from the domain of standard usability evaluation methods which are used by evaluators. The expression "generic method" is used in order to distinguish such a method from a "commercial method" which offers a composite solution and for which licences and fees may be payable.

OVERVIEW OF USABILITY EVALUATION
Throughout specialist text books, journals and standards, there is an on-going confusing usage of terms like model, method, framework, technique and tool (Dix *et al*. 1998; Holcomb and Tharp 1991; ISO/DIS 9241-11 1995; Whitefield *et al*. 1991). This paper shows that, during software usability evaluation, the correct usage of terms starts with "strategies" for evaluating usability and that these "strategies" employ a variety of evaluation "methods". A two-by-two matrix of strategies for evaluating software usability is presented in this paper.

According to ISO 13407 (1997) the purpose of usability evaluation is

1. To provide feedback to improve design.
2. To assess that user and organisational objectives are being achieved.
3. To monitor long term use of product or system.

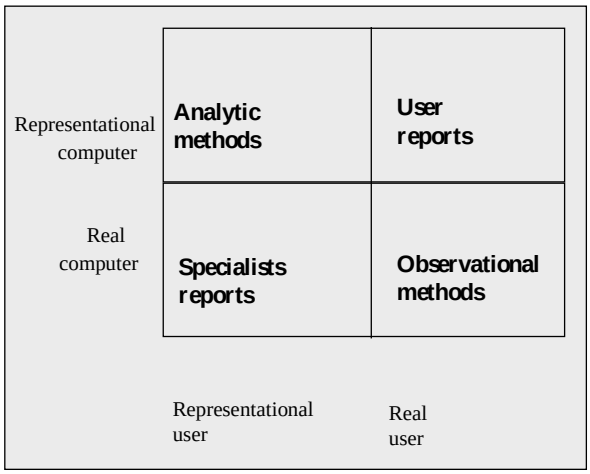
Wixon and Wilson (1997) offer a broader view and suggest that usability engineering puts the user at the centre of the process. They point out that for the engineering community usability engineering provides them with an understanding of the user's viewpoint of a system. As part of cost/benefit analysis, it allows the financial community to quantify the benefits of a system. It allows the marketing community to justify the quality features of a system in "objective and holistic terms". And, for buyers it is a means of evaluating a system in terms of their usability requirements (Wixon and Wilson, 1997).

A usability evaluation method is a systematic procedure for recording data relating to end-user interaction with a software product or system. This recorded data can then be analysed and evaluated in order to determine the usability, or otherwise, of the product. There are a number of different generic methods that can be used during a usability evaluation session. Depending on the cost and life cycle considerations methods are used to secure feedback which will improve usability or they can be used to establish whether usability is "good enough" (i.e. meets some criterion). Unfortunately, there is no universally accepted set of names for the methods. While most authors (Kirakowski 1995; Nielsen 1993; Preece *et al*. 1994) consistently use some method names (e.g. observation, questionnaire, interview, heuristics), other method names (e.g. cognitive walkthroughs, predictive and interpretative evaluation) used by authors are a reflection of the individual author's bias or preference (Dix *et al*. 1998; Preece *et al*. 1994). In order to clarify the domain of methods that exist, this paper uses three typical sets of methods by established and recognised authors (Dix *et al*. 1998; Nielsen 1993; Preece *et al*. 1994) in the domain of usability evaluation as the basis for identifying generic methods. These methods are then classified using a two-by-two matrix of strategies and presented as a taxonomy of generic methods. A Usage Analysis Table (UAT) of the various methods is also presented.

STRATEGIES FOR EVALUATING SOFTWARE USABILITY

Whitefield *et al*. (1991) proposed a classification of usability evaluation methods and this section makes use of their core thinking to clarify the terms that will be used in the remaining sections of this paper. Whitefield *et al*. sub-divide the resources that are available during evaluation into four categories - Real users, Real computers, Representational users and Representational computers - and they position these categories on a two-by-two matrix. They then explain that usability evaluation methods can be classified to suit this matrix and proceed to classify the methods as *Observational methods, Specialists reports, User reports and Analytic methods.* Whitefield's classification diagram is shown in figure 1.

Figure 1 - *Classes of Evaluation Methods by Whitefield et al.*



In figure 2, the quadrants have been inverted to better reflect the high/low measure associated with two-by-two matrix diagrams. This is done on the basis that it is generally accepted that the best feedback and results are obtained in the real/real scenario (Holcomb and Tharp 1991).

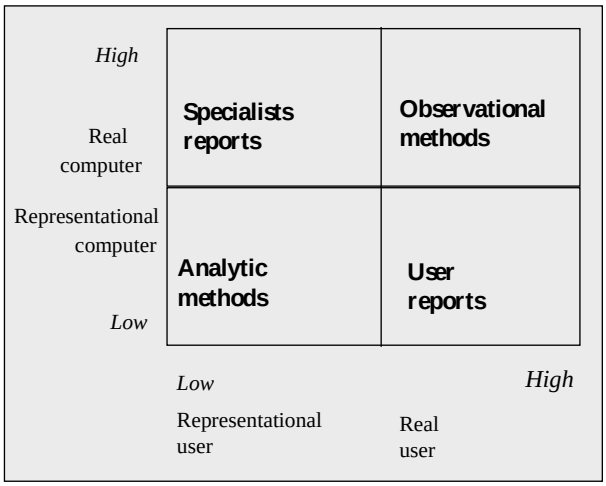


Figure 2 - *Adapted Layout*

There are problems with Whitefield's approach because their classification implies that:

1. Only observational methods can be used with real user and real computers. This is not the case as when real computer resources and real users are available, then a

number of usability evaluation methods can and should be used.

2. User reports can only be obtained when real users and representational computers are available. This is equally untrue because user reports can be obtained from both of the real user quadrants. Furthermore, it is desirable that they should be.
3. Specialists reports can only be obtained from representational users, who by definition are "descriptions or models of users". This would not be in keeping with a natural understanding of specialist. The word specialist in specialists reports implies something different to the defined real or representational user. Therefore, a specialist or expert report can be returned from all four quadrants.
4. It is appropriate to comment on previous approaches to classifying evaluation methods. Dix *et al*. (1998), who prefer the word technique to method, also seek to classify usability evaluation methods but comment that their *classification is intended as a rough guide only - some of the techniques do not fit easily into such a classification since their use can vary considerably".*

In order to overcome the difficulties explained in the last paragraph, it is necessary to take into account the stage in the lifecycle when usability evaluation takes place (Dix *et al*. 1998). It is also necessary to take into account the desirability of employing multiple methods during the evaluation process. So, this paper suggests that an appropriate "strategy" be used which reflects the lifecycle timing. The strategy used will also depend on the resources that are available. For example, when real users and real computers are available, the usability evaluation strategy can involve real users, doing real tasks, using real equipment, in a real environment. This paper calls this strategy ***Real World***. A completely different strategy is necessary when both the user and the computer are representational. As suggested by Whitefield *et al.* (1991), analytic methods have to be employed which are easily conceptualised as being the opposite to the real world. To reflect that it is diametrically opposite to the real world and to reflect the engineering aspect of the activities that have to be performed, this paper calls this strategy ***Virtual Engineering***. The third strategy involves real users and representational computers. In this domain, soft, is well understood as relating to users and user involvement and because modelling is a substantial part of this strategy, this paper calls this strategy ***Soft Modelling***. Finally, the fourth strategy involves representational users (including experts who can conduct critical reviews) with real computers and software product. This paper calls this strategy ***Hard Review***. This classification of strategies is presented in figure 3.

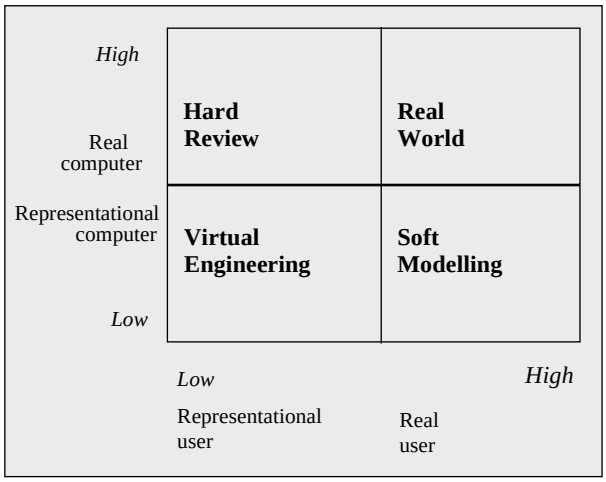


Figure 3 *Strategies for Usability evaluation*

The methods to be used and the timing of evaluation during the lifecycle are important considerations for strategic managers and IS professional. By using meaningful names and by addressing the system lifecycle, the strategies outlined provide evaluators with an enhanced perspective of these important considerations. Furthermore, the expression strategy reinforces that there are multiple personnel resources involved. The personnel resources include evaluators, users, developers, financial and marketing staff. The term strategy also reinforces the need for planning in order to ensure the most efficient use of funding and technological resources.

The remainder of this paper focuses on the domain of generic usability evaluation methods that can be used in these strategies.

METHODS FOR EVALUATING SOFTWARE USABILITY

The four strategies provide a framework for HCI professionals and strategic managers who are responsible for software usability evaluation. Within each strategy a number of different usability evaluation methods can be employed and in this section these methods are identified. To accurately identify which usability evaluation methods are used, the methods explained in three popular and current texts written by acknowledged authors (Dix *et al*. 1998; Nielsen 1993; Preece *et al*. 1994) are considered. These are important texts which are used extensively with third-level syllabus and are therefore well known to students and practitioners. The methods are presented in figure 4. Readers who require further detailed explanation of each evaluation method are referred to the explanations in the three selected texts.

Dix et al. (1998)	Nielsen (1993)	Preece et al. (1994)	Composite list
Observation	Observation	Observing and monitoring usage	• Observation
Query techniques (Interviews and questionnaires)	Questionnaires and interviews	Collecting user opinions (Interviews and questionnaires)	• Questionnaire • Interview
Empirical methods		Experiments or benchmark tests	• Empirical Methods
	Focus groups		
	Logging actual use		• User Groups
	User feedback		
Cognitive walkthroughs			• Cognitive Walkthroughs
Heuristic evaluation	Heuristic evaluation		• Heuristic Methods
Review based evaluation			• Review Methods
Model based evaluation			• Model Methods
		Predictive Evaluation	
		Interpretative Evaluation	

Figure 4 - *Comparison and Composite list of Usability Evaluation Methods*

From this selection of popular generic usability evaluation methods it is possible to prepare one composite list of methods. The composite list that this paper proposes and which will be used in the next section is set out in the fourth column of figure 4. A short general description of each method is given in figure 6.

Preece *et al*. (1994) include two groupings of evaluation methods which they name as Predictive and Interpretative evaluation. Their Predictive methods are similar to Cognitive walkthroughs, Heuristic Methods and Model Methods. Their Interpretative evaluation includes methods like ethnography and other user focused methods. So, for the purpose of this paper both of these groupings are considered to be included in the composite list.

SELECTING AN APPROPRIATE USABILITY EVALUATION METHOD

A primary consideration for the professional - who is trying to evaluate the usability of a computer system - is how to select appropriate usability evaluation methods. The four different usability evaluation strategies tend to mirror the phases of the system development lifecycle (see Sommerville 1992 for a description of the software lifecycle). The lifecycle progresses from analysis and design through prototyping and development to installation. Roughly corresponding with these phases are the strategies Virtual Engineering at design stage through Soft Modelling and Hard Review to Real World at the installation stage. Some usability evaluation methods are inappropriate for use in individual strategies. For example, observation is inappropriate in a Virtual Engineering strategy because there is no real user to observe. On the other hand, some methods are fundamental to particular strategies. In any particular strategy, the evaluator should seek to use more than one usability evaluation method (this is desirable and in some instances necessary) to secure the most reliable result. This view is supported by Reiterer and Oppermann (1993), who explain that:

"there is no "single" best evaluation method. All of the methods have some disadvantages, or consider only a limited number of the factors influencing an evaluation, but many of them contain useful ideas, or are very appropriate for the evaluation of a specific factor. What is needed is a combination of different evaluation methods for the different foci of an evaluation".

	Observation	Questionnaire	Interview	Empirical methods	User groups	Cognitive walkthroughs	Heuristic methods	Review methods	Model methods
STRATEGY Real World Hard Review Soft Modelling V. Engineering	R.W. S.M.	R.W. S.M.	R.W. S.M.	R.W. S.M.	R.W. S.M.	 V.Eng	 H.R. V.Eng	 V.Eng	 V.Eng
LOCATION Laboratory Work context (Field)	Lab. Field	Lab. Field	Lab. Field	Lab.	Lab. Field	Lab.	Lab.	Lab.	Lab.
BIAS Objective Subjective	Objective Subjective	Objective Subjective	 Subjective	Objective	Objective Subjective	 Subjective	 Subjective	 As source	 Subjective
USABILITY MEASURE Effectiveness Efficiency Satisfaction	Yes Yes Yes	Yes Yes Yes	Yes Yes Yes	Yes Yes Yes	 Yes	Yes Yes	 Yes	 As source	Yes Yes Yes
INFORMATION High level Low level	High Low	High	High	High Low	High Low	 Low	High Low	 As source	 Low
IMMEDIACY OF RESPONSE Immediate Delayed	Immediate Delayed	 Delayed	Immediate	Immediate Delayed	Immediate Delayed	Immediate Delayed	Immediate	 As source	 Delayed
INTRUSIVE	Yes	No	Yes	Yes		No	No	No	No
COST Expensive Inexpensive	Audio rec. Video rec. Comp. logs Post-task Paper & pen Think aloud User note-books	Inexpensive.	Inexp.	Exp.	Inexp.	Inexp.	Inexp.	Inexp.	Inexp.

Figure 5 - *Usage Analysis Table. Appropriate use of usability evaluation methods - after Dix et al.*

General descriptions of usability evaluation methods	
Generic method	**General description**
Observation	A usability evaluation specialist acts as the observer of users as they interact with computers, noting user successes, difficulties, likes, dislikes, preferences and attitudes.
Questionnaire	The use of a set of items (questions or statements) to capture statistical data relating to user profiles, skills, experience, requirements, opinions, preferences and attitudes.
Interview	A formal consultation or meeting between a usability evaluation specialist and user(s) to obtain information about work practices, requirements, opinions, preferences and attitudes.
Empirical Methods	The testing of a well defined *hypothesis* by measuring *subject* (user) behaviour while the evaluator manipulates *variables*.
User Groups	Availing of the wealth of knowledge and experience of organised (user forum) and selected (beta site) end users.
Cognitive Walkthroughs	A step by step evaluation of a design by a cognitive psychologist in order to identify potential user psychological difficulties with the system.
Heuristic Methods	The use of a team of usability evaluation specialists to review a product or prototype in order to confirm its compliance with recognised usability principles and practice.
Review Methods	The review and reuse of the wealth of experimental and empirical evidence in the research literature and in the de-facto standards established by the software industry.
Modelling Methods	The use of models like GOMS (Goals, Operations, Methods and Selection) and KLM (Keystroke Level Modelling) to predict and provide feedback on user interactions and difficulties.

Figure 6 - *General descriptions of usability evaluation methods*

Matching the strategies to software engineering lifecycle is an important step which in turn makes it possible to identify which methods should be used in combination when evaluating usability.

Usage Analysis of Usability Evaluation Methods

When selecting an appropriate usability evaluation method or combination of methods, the selector will need to take into consideration the different foci of the evaluation. Dix *et al*. (1998) suggest that these foci or considerations are:

- The stage in the lifecycle at which the evaluation is carried out.
- The style of the evaluation.
- The level of subjectivity or objectivity of the method.
- The type of measures provided.
- The information provided.
- The immediacy of the response.
- The level of interference implied.
- The resources required.

Dix *et al.* (1998) describe these considerations as factors which distinguish evaluation techniques (Dix *et al.* use the term technique in preference to method). These factors together with the methods in the composite list of usability evaluation methods can be combined and a summary analysis of each method presented. This analysis provides information for HCI professionals who have organisational responsibility for ensuring the usability of software used within their organisation. This approach is based on a set of similar classification tables used by Dix *et al.* which combines their list of techniques, i.e. methods, and their set of factors. Their set of factors are named *Stage, Style, Objective/Subjective, Measure, Information, Immediacy, Intrusive, Time, Equipment and Expertise.* However, in this paper the factors have been adapted to incorporate the four strategies for usability evaluation as shown in figure 3. Measure has been adapted to reflect the usability measures of effectiveness, efficiency and satisfaction as required by ISO/DIS 9241-11 (1995) and a separate cost consideration is shown in this paper. The result is shown in figure 5 and is referred to as the "Usage Analysis Table.

Explanation of Usage Analysis Table

This subsection explains in turn each of the considerations set out in figure 5.

Strategy

The first consideration is the stage in the lifecycle at which the evaluation is performed. This maps to the four strategies and is shown as the first entry on figure 5. An analysis of this entry will be presented later as a taxonomy of methods.

Location

The second consideration is the location of the evaluation which refers to whether it is done in a laboratory or in a real work environment.

Bias

The third consideration is the level of subjectivity or objectivity inherent in the method. The results and feedback from some methods can be greatly influenced by the bias or preferences of the specialist or facilitating evaluator. Observation is an objective method but is shown as objective and subjective. This is because this

method can involve elements of evaluator- and user-input which both add a subjective aspect to the results.

Usability Measure
The next consideration is the type of measures provided by the method. Dix *et al.* (1998) describe these as quantitative or qualitative, explaining that quantitative are usually numeric based and can be easily analysed using statistical techniques. Qualitative are non-numeric and relate to user preferences and attitudes. The usability measures suggested in ISO/DIS 9241-11 (1995) and Bevan and Macleod (1994) are effectiveness, efficiency and satisfaction. So, Usability Measure is subdivided into effectiveness, efficiency and satisfaction to better reflect the wording of ISO/DIS 9241-11 (1995).

Information
The fifth consideration is the information or feedback provided by the method. Information is described by Dix *et al.* (1998) as low-level, which, for example, would give feedback on topics like most readable font, most appropriate colour combinations, easiest recognised icon and similar design information. High-level feedback is more of an overall impression of the general usability of the system. As can be seen from figure 5, some methods return both high and low level information.

Immediacy of Response
Immediacy of response is concerned with how quickly the feedback is available. For example, during observation, video recordings have to be reviewed and analysed which can take a considerable time. So, results are delayed for this situation. However, paper & pen records are available as soon as the observation is concluded which means that the results are immediate. Also during observation, post-task walkthroughs have to be left for analysis until the user is no longer engaged at the workplace. Time must also be scheduled for reviewing the videos, so these results will be delayed. On the other hand, a good interviewer can come away from an interview session with clear criteria to be avoided and/or included in an interface.

Intrusive
Intrusiveness is concerned with the fact that some users behaviour might be influenced by the presence of an observer, interviewer or recording system to the extent that the results of the evaluation session might not be accurate.

Cost
The final consideration in Dix's list is Resources. This is not shown separately in figure 5 because the specific user (human resource) and computer resource are already dictated as part of the chosen strategy (Real user, Real computer, Representational user and Representational computer). However, information systems professionals and business managers are concerned with costs and a simple Expensive/Inexpensive indication is included.

The Usage Analysis Table (UAT) owes its concept to tables in Dix *et al.* However, there are four key differences.

1. The UAT is based on an extended list of evaluation methods and it presents information in one complete table (instead of three).
2. The Strategy element reinforces the software engineering lifecycle and identifies which of the extended list of methods are appropriate during the different stages of the lifecycle. This entry reflects a significant issue which is of interest to current researchers and practitioners (Bevan and Azuma, 1997;IN USE, 1997; Karat, 1997)
3. The recommended measures from ISO/DIS 9241-11 (1995) are used on the table. This is a significant element in the UAT and provides an excellent analysis of what each method can evaluate.
4. Cost is another significant issue which is of interest to current researchers and practitioners (IN USE, 1997; macleod *et al*., 1997: Wixon and Wilson, 1997). It is addressed by Nielsen (1993) and is a motivator for Discount Usability. The UAT highlights cost across all methods which in turn is reflected in the strategies and consequently can be budgeted for in cashflow forecasts.

TAXONOMY OF USABILITY EVALUATION METHODS

In keeping with the view that more than one usability evaluation method is desirable and in some instances necessary to secure the most reliable results, it is now appropriate to combine the enhanced two-by-two matrix, the Composite list of methods and the Usage Analysis Table. The strategy focus of figure 3 (after Whitefield *et al*.) is combined with the usage analysis of figure 5 and is converted into a taxonomy of usability evaluation methods. This is presented in figure 7. It shows the different usability evaluation methods (from the composite list shown in figure 4) properly positioned on the two-by-two matrix of usability evaluation strategies shown in figure 3.

HCI professionals who need to select methods for usability evaluation can consult the Strategies (figure 3), the Usage Analysis Table (figure 5) and the Taxonomy of Usability Evaluation Methods (figure 7) to help them decide which methods are appropriate for their evaluations. They can be confident that a selected method will address their specific concerns. The Usage Analysis Table can also be used to support management when making decisions regarding the allocation of funding for evaluation.

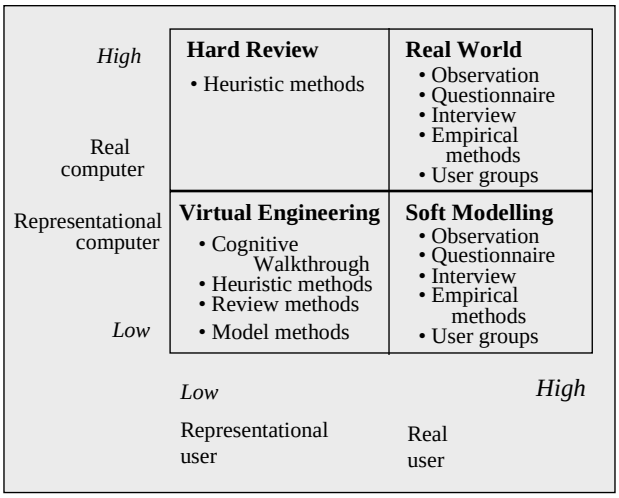


Figure 7 - *Taxonomy of Usability Evaluation Methods*

CONCLUSION

The usability evaluation of computer systems is an important issue for strategic managers and IS professional. In order to assist such professionals in the selection and understanding of the various popular usability evaluation methods, this paper presents a methodical analysis and critical review of the difficulties and confusion identified by reliable HCI researchers and presents an alternative and significantly enhanced view which resolves these difficulties. In particular this paper has

- outlined a two-by-two matrix of strategies for evaluating software usability,
- identified the domain of generic methods that are used for usability evaluation,
- presented a Usage Analysis Table which illustrates appropriate usage of each generic method and incorporates latest International Organisation for Standardisation guidelines for usability evaluation, and cost,
- presented method classification as a taxonomy of methods.

The paper adds additional perspective to our understanding of usability evaluation by emphasising lifecycle timing and costing. In particular the paper emphasises the importance of a strategic approach to usability evaluation. These tools can be used by HCI and Information Systems professionals as part of their usability evaluation activities.

REFERENCES

1. Bevan, N. and Azuma, M. (1997) Quality in use: Incorporating human factors into the software engineering lifecycle. In: Proceedings of the Third International Software Engineering Standards Symposium and Forum (ISESS'97), p169-179
2. Bevan, N. and Macleod, M. (1994) Usability measurement in context, Behaviour and Information Technology, 13, p132-145
3. Dix, A., Finlay, J., Abowd, G. and Beale, R. (1998) *Human-computer interaction*, Prentice Hall, Hemel Hempstead, UK
4. Holcomb, R. and Tharp, A. (1991) "Users, a software usability model and product evaluation", *Interacting with Computers*, Butterworth-Heinemann, Oxford, UK, Vol 3(2) pp. 155-166
5. ISO/DIS 9241-11 (1995) *Draft International Standard. Ergonomic requirements for office work with visual display terminals (VDTs). Part 11:Guidance on usability*, International Organisation for standardisation, Genève, Switzerland
6. ISO/DIS 13407 (1997) *International Standard - Human-centred design processes for interactive systems*, International Organisation for Standardisation, Genève, Switzerland
7. Karat, J. (1997) User-centred software evaluation methodologies, Handbook of HCI, 2nd edition, Elsevier Science, p689-704
8. Kirakowski, J. (1995) "The Software usability measurement inventory: background and usage", *Unpublished collection of readings*, Human Factors Research Group, University College Cork, Ireland (To appear in Jordan, P., Thomas, B. and Weerdmeester, B., *Usability evaluation in industry*, Taylor and Francis)
9. Macleod, M., Bowden, R. and Curson, I. (1997) The MUSiC performance measurement method, Behaviour and Information Technology, 16
10. Nielsen, J. (1993) *Usability engineering*, Academic Press Limited, London, UK
11. Preece, J., Rogers, Y., Sharpe, H., Benyon, D., Holland, S. and Carey, T. (1994) *Human-computer interaction*, Addison-Wesley, Wokingham, UK
12. Reiterer, H. and Oppermann, R, (1993) "Evaluation of user interfaces:EVADIS II -a comprehensive evaluation approach", *Behaviour and information technology*, Taylor and Francis, Basingstoke, UK, Vol 12(3), pp. 137-148
13. Sommerville, I. (1992) *Software engineering*, Addison-Wesley, Working, UK
14. Whitefield, A., Wilson, F. and Dowell, J. (1991) "A framework for human factors evaluation", *Behaviour and Information Technology*, Taylor & Francis Ltd, Basingstoke, UK, Vol 10(1), pp. 65-79
15. Wixon, D. and Wilson, C. (1997) The usability engineering framework for product design and evaluation, Handbook of HCI, 2nd edition, Elsevier Science, p653-688