

1999-01-01

A process for appraising commercial usability evaluation methods.

Ronan Fitzpatrick

Technological University Dublin, Ronan.fitzpatrick@tudublin.ie

Alan Dix

Technological University Dublin

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomcon>



Part of the [Computer Engineering Commons](#)

Recommended Citation

Fitzpatrick, Ronan and Dix, Alan: A process for appraising commercial usability evaluation methods. Human-Computer Interaction: Ergonomics and User Interfaces, Proceedings of HCI International '99, Munich. Published by Lawrence Erlbaum Associates.

This Article is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)

A Process for Appraising Commercial Usability Evaluation Methods

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

Alan Dix

School of Computing,
Staffordshire University,

Beaconside, Stafford ST18 ODG,
United Kingdom.

Tel: +44 1785 353428, Fax: +44 1785 353431

Email: A.J.Dix@soc.staffs.ac.uk

ABSTRACT

Recent international quality standards and European Community legislation have identified new software quality factors. These new factors include *suitability*, *installability* and *adaptability*. Other quality factors need to be reviewed in the light of these developments. This has impacted on established commercial usability evaluation methods to the extent that it is appropriate to ask if these evaluation methods comply with the new standards and legislation. In order to answer this question the commercial evaluation methods need to be appraised (meta-evaluation) using a suitable method appraisal process. This paper describes such an appraisal process which specifically addresses the many considerations raised by the standards and the law. The appraisal method consists of two parts which provide an overview of the commercial method and a methodical analysis of how it complies. By combining this analysis with a weighting and rating technique the appraised method can achieve a score which can be compared with other commercial methods. The process is an essential tool for strategic managers who are responsible for usability evaluation during systems acquisition. It is also of benefit to supplier organisations who, in their efforts to develop the highest quality systems, need to demonstrate compliance with international standards and development process maturity models.

Key words

International standards, European Community legislation, software quality, usability evaluation, commercial evaluation methods, strategic application, life cycle processes, context of use, usability measures, usability attributes.

INTRODUCTION

For the purpose of supporting software usability evaluation, a number of commercial usability evaluation methods have been developed by academics and industry. A commercial usability evaluation method is a practical implementation of a generic method and incorporates good professional practice and industry standards. Most of these methods predate ISO 9000-3 [7], ISO/DIS 9241-11 [8] and the European Community Council Directive on display screen equipment [3]. Consequently, it is necessary for evaluating organisations to appraise these commercial methods in the light of these new standards and legal obligations. This paper defines a process for this appraisal. The paper starts with a clear explanation of the need for tools that support information systems professionals and strategic managers who have organisational responsibility for usability evaluation. In particular, it explains the need for a process for appraising commercial usability evaluation methods and the issues that should be addressed by an appraisal process are clarified. It continues by defining the appraisal process using A Method Appraisal Grid (MAG). Each entity on the Grid is fully explained. These entities include Methodology, Strategic application, Life cycle processes, Context of use, Usability measures, Attributes of a usable software product, method scoring (Weighting and Rating) and Method Reliability. The advantages of the Method Appraisal Grid are also reviewed. The paper explains how the process is applied and illustrates its application to two sample usability evaluation methods. Finally the paper suggests a number of additional beneficial uses of the MAG.

RATIONALE

The current need for systems professionals and strategic managers to comply with international quality standards and statutory obligations has placed additional emphases on the importance of usability evaluation [21], [22]. There is a need to know that systems (new systems being developed, systems purchased off-the-shelf or legacy systems being re-engineered) comply with the most current usability standards and with current legal

recording of its features and usefulness to the usability evaluation process. The second stage of the appraisal process involves identifying the usability considerations that are addressed by the method. A weighting and rating matrix is then used for the purpose of scoring the method. The results of the process are documented on a specially designed form called the Method Appraisal Grid. This Grid is shown in figure 1. This form is specifically designed to address the need-to-know issues outlined previously. It is divided into two parts - reflecting the two stage approach. In the first part, headed Usability evaluation method, the form provides space for each method to be named and for recording its Methodology, Strategic Application and support for Life cycle processes. The second part of the form, which is headed Usability considerations, is used to indicate "what" the method evaluates. This includes Context of use, Usability measures and the Attributes of a usable software product. [2], [8], [5]

Support for a weighting and rating scoring system is incorporated for completion by the appraiser. Individual elements are weighted to suit the evaluating organisations specific requirements and the extent to which these requirements are met is rated. The weighting and rating process is explained in greater detail in later sections. From this, a score is calculated for each commercial method being reviewed. Finally there is space to record a proven and accepted reliability measure for each method (as published by the method owner or sponsor). The content of the Method Appraisal Grid is now fully explained in the following sub-sections.

Methodology

The Methodology category is included to allow systems professionals and business managers to see at a glance the motivating philosophy for the method, if it is a model or a method and if it is supported by appropriate tools. They can see if it is a standalone method or if it is part of an integrated suite. The sub-divisions, *Philosophy*, *Model*, *Generic Method* and *Tool*, are based on the view of methodologies given by [1]. Readers should refer to Avison and Fitzgerald's work for a full explanation of these terms. Particularly useful, for example, is an indication on the Grid of whether or not the method being appraised is supported by some form of a computerised tool.

Strategic Application

The Strategic Application of the commercial method is based on four strategies which take into account the stage in the life cycle when usability evaluation takes place. It is also necessary to take into account the desirability of employing multiple methods during the evaluation process [21], [23]. The strategy used will depend on the resources that are available [11], [17], [23]. 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. In this strategy analytic methods are employed [23]. These 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*. These strategies are classified in figure 2.

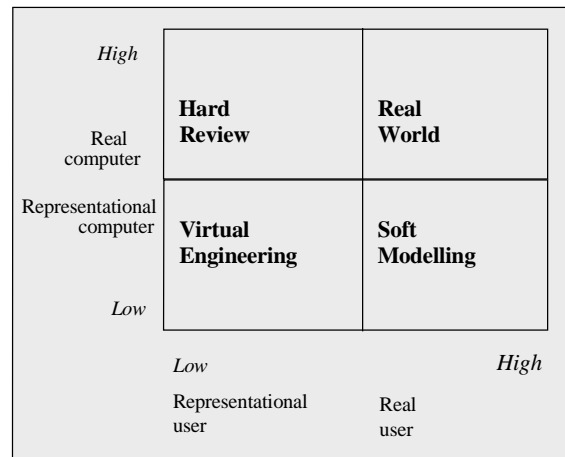


Figure 2 - Strategies for Usability Evaluation

How an appraiser should record an entry in this Strategic Application category on MAG will be illustrated later.

Life cycle processes

The third category on the Grid records which system life cycle stages or processes the method can be applied to or is limited to. From this category, it is easy to see "where" in the life cycle of a system the evaluation method under review can be used. This category is "loosely" sub-divided on the basis of the primary processes and activities outlined in [10]. Sometimes the software to be evaluated will be a generic product which can be bought off-the-shelf. In this case there are limited development stages or processes. So, it is appropriate to establish if the commercial method offers support for those who wish to evaluate product they buy rather than evaluate product they build. The life cycle sub-divisions which are used are *Analysis*, *Design*, *Build*, *Buy*, *Operation* and *Maintenance*.

Context of use

Context of use is derived from [8] and [9]. According to these standards, when evaluating usability it is necessary for evaluators to consider the users profile, the tasks that users will be completing, the equipment to be used and the environment where the user will work. So, this category is used to identify whether the method being reviewed addresses these context of use topics. The headings are *User*, *Goals/Tasks*, *Equipment* and *Environment*. A detailed explanation of context in use is given by [2].

Usability measures

The Usability measures used on the Grid are those recommended in ISO/DIS 9241-11 and are *Effectiveness*, *Efficiency* and *Satisfaction* [8]. In addition, empty rows are included on the Grid so that the appraisal can be tailored by the appraiser to suit special usability measures if needed. For example, *Usage* [6] might be another measure which is of interest to some appraisers.

Attributes of a usable software product

The Attributes of a usable software product are based on the external software quality factors contained in a listing developed by [15] combined with a list of factors derived from a methodical analysis of factors contained in ISO 9000-3 and in the European Community Council Directive relating to display screen equipment - [3], [5], [7].

Each of the Attributes of a usable software product must be reviewed in order to establish if the commercial method addresses the attribute. So, all of the attributes listed by [5] are included on the Grid. These are, *Suitability*, *Installability*, *Functionality*, *Adaptability*, *Ease-of-use*, *Learnability*, *Interoperability*, *Reliability*, *Safety*, *Security*, *Correctness*, *Efficiency* and an *Overall* product attribute. As part of the evaluation of user-satisfaction it is necessary to establish satisfaction with the "overall" software product [12]. So, "overall" is considered to be another attribute of a usable software product and is, therefore, added to the Grid for this appraisal.

Weighting

Both the Context of use and the Attributes of a usable software product are included in the scoring process. The process uses weighting factors to represent the importance of the sub-sections of the context of use and the attributes of a usable software product. Because each organisation will have different usability needs, the weighting factors are left to the organisation's own decision. For example, consider an organisation that has a broad range of end-users who vary from novice to "power users". In this case, being able to adapt the interface to suit end-user needs might be more important than installability might be. Consequently, *Adaptability* would be weighted higher for this appraisal than *Installability*. Or, if the system was a

walk-up-and-use one in a bank which allowed customers to view their bank statements, then *Adaptability* might not be a consideration at all, while *Ease-of-use* and *Reliability* are.

Consequently, *Ease-of-use* and *Reliability* would have a higher weighting than *Adaptability* or *Installability*. So, before commencing an appraisal of a commercial method, the evaluating organisation must decide the relative importance of context of use and software attributes and then ascertain a percent value for each, such that the total of the weightings add up to 100%. Weighting is only shown once on the Grid because the same weighting must be consistently applied by the organisation to all methods being appraised.

Rating

During the rating process the appraiser examines each component of Context of use and each Attribute of a usable software product to establish the extent to which the method being appraised addresses these considerations. Based on this examination a rating value is applied to each consideration. The rating is the subjective grading on a scale of 0 to 100 of the extent to which the different components of contexts of use and attributes are, in the opinion of the appraiser, satisfied by the method under review. For example, consider that it is intended to acquire software for use in a workgroup situation, and because of the importance of *Interoperability* to this situation, this attribute would be given a high weighting. Now, if the method being appraised is a questionnaire for measuring user-satisfaction and if it asks no questions about *Interoperability*, then the appraiser would rate this attribute at 0 (out of 100). Or, if the appraiser was examining *Suitability* and was only partially satisfied with how the method evaluated *Suitability*, then the appraiser would rate the attribute accordingly (say 60 out of 100, if 60% satisfied). Using a similar subjective approach, the appraiser rates (on the scale of 0 to 100) all components of the Context of use and the Attributes of a usable software product. The weighting and rating are then used to calculate a score or usability quotient [5] for the method being appraised.

Scoring

A method is scored by multiplying each weighting by its corresponding rating and adding all the quotients [22]. This gives a score that can be used for method comparison.

Method Reliability

System professionals need to be confident that the results yielded by the chosen commercial method are reliable [12], [21] So, the final record on the Method Appraisal Grid is Method Reliability.

under the Buy heading would be a suitable entry in this category for the EVADIS II method because that is a recommended usage [21]. Alternatively a simple "Yes" under any heading or any other description preferred by the evaluator will indicate appropriate usage.

Context of use

Context of use uses short phrases relating to the four components in this category. Typical suitable phrases are:

- Novice/expert/age <30
- Per ISO 9241-11
- Per EU Directive
- Not considered

Evaluators will devise their own phrases, as appropriate, for their own organisational needs.

Attributes of a usable software product

Each of the Attributes of a usable software product is reviewed in turn to establish if the usability measures (*Effectiveness*, *Efficiency*, *Satisfaction*) are addressed by the commercial method. A Yes/No or Y/N notation is used. The appraiser should start with *Effectiveness* and ask, "Does the method address user-effectiveness in achieving *suitability*?" followed by "Does the method address user-effectiveness in achieving *functionality*?", and so on until all attributes have been reviewed. However, to ensure that user-effectiveness is properly understood, its definition from an international standard is recalled. It is, "the accuracy and completeness with which users achieve specified goals" [8]. Applying this definition or interpretation, the appraiser would ask, "Does this method address

- user accuracy and completeness in achieving *suitability* goals?
- user accuracy and completeness in achieving *installability* goals?
- user accuracy and completeness in achieving *functionality* goals?
- user accuracy and completeness in achieving *adaptability* goals?
- user accuracy and completeness in achieving *ease-of-use* goals?
- user accuracy and completeness in achieving *learnability* goals?
- user accuracy and completeness in achieving *interoperability* goals?
- user accuracy and completeness in achieving *reliability* goals?
- user accuracy and completeness in achieving *safety* goals?
- user accuracy and completeness in achieving *security* goals?
- user accuracy and completeness in achieving *accuracy* goals?
- user accuracy and completeness in achieving *efficiency* goals?

- user accuracy and completeness in achieving *overall* goals?"

Reviewing these questions it becomes obvious that the effectiveness measure will not apply to all of the attributes. *Installability*, *Functionality*, *Adaptability*, *Ease-of-use*, and *Learnability* might be considered as primary user-effectiveness issues because it is the user who has to install, it is the user whose effectiveness is dependent on functionality, it is the user who has to adapt the interface, it is the user who needs an easy to use interface and it is the user who has to learn to use the product. However, the relevance of the effectiveness of the remaining attributes is for the appraiser to decide in the best interest of the organisation.

In the *Methodology* category, the illustrated example (Sample 1) suggests that the method might be a questionnaire. If this is the case then the appraiser will need to be satisfied that sufficient items (questions) are included in the questionnaire to adequately evaluate each attribute. For example, if there are no questions with an *Adaptability* focus, then, does the questionnaire evaluate *Adaptability*?

The second usability measure is *Efficiency* and it is defined as the "resources expended in relation to the accuracy and completeness with which users achieve goals" [8]. *Efficiency* is subdivided into human, temporal (time) and economic (cost). *Efficiency* is shown in the international standard to be a function of *Effectiveness*, so, *Efficiency* applies to the same attributes as *Effectiveness*.

For the *Satisfaction* measure, each attribute is reviewed to establish if the method being appraised addresses user-satisfaction with regard to that attribute. So, the appraiser will start with *Suitability* and ask "Does this commercial method evaluate user-satisfaction with regard to *Suitability*?" The appraiser will record a "Yes" or "No" as appropriate and repeat the procedure for each attribute in turn. It is obvious that this measure applies to all attributes.

Scoring

The second part of the Grid incorporates a facility for scoring the different usability evaluation methods using a weighting and rating matrix.

To illustrate the scoring system the Method Appraisal Grid has been completed using two Sample methods - Sample 1 and Sample 2. For the purpose of this illustration, the four components of *Context of use* and all but one of the *Attributes of a usable software product* are equally weighted at 5%. *Safety* is weighted at 20% to reflect a special importance attached to it in this illustrative appraisal. These percent values (5% and 20%) are solely for illustrative purposes and they bear no relation to any

specific software product. The appraiser considers each of the usability measures in turn and records a "Yes" or "No" for each attribute which indicates if the method addresses these attributes. The results are then used to subjectively rate the extent to which the method addresses these attributes using a 0 to 100 scale. Where the method does not address a particular component of context of use or attribute then that element is rated at 0. All of the values obtained by multiplying individual weights by their corresponding rates are added to give a score for the Sample 1 method.

For comparison purposes, the appraisal of a second (Sample 2) method is also illustrated. In this case all of the usability considerations are considered to be fully addressed by the method, so a "Yes" is recorded for each. Furthermore, each element fully addresses the appraiser's needs so they are all rated at 100. This gives a total score for the Sample 2 method of 100.

Method Reliability

The final category, Method Reliability, provides space for to record a numeric value for the reliability of the method as published by the method owner or an independent assessors verification of that claim.

USING THE METHOD APPRAISAL GRID

There are a number of uses for the Method Appraisal Grid.

- Its primary use is to appraise selected commercial evaluation methods where a specific off-the-shelf product or a build project has been identified.
- Where it is necessary for the evaluating team to devise their own or customise an existing commercial method, MAG provides a focus for the essential categories that must be addressed.
- Usability evaluation is essential to the quality control of a project. But, this evaluation is only useful if it assess the appropriate usability factors for the software project. The MAG allows managers to choose the correct usability methods for their project thus ensuring the highest quality software. Consequently, supplier organisations seeking to show compliance with a quality development process like the Capability Maturity Model [18] can use MAG to demonstrate the quality-focus of their evaluation practice.
- MAG is of benefit to developer organisations who wish to demonstrate compliance with ISO 13407 (Table A5) regarding evaluation of designs against user requirements. To comply with the requirements of the standard, these developers will be able to cite MAG as "The process and rationale for the selection of [usability] methods and measures used".

CONCLUSION

This paper has defined a process for appraising the usefulness of commercial methods that are used for

evaluating software usability. For the purpose of documenting the results of such an appraisal, a Method Appraisal Grid (MAG) has been devised. This Method Appraisal Grid incorporates guidelines and statutory obligations set out in international standards and European Community legislation. The content of the Grid was fully explained and two illustrative methods (Sample 1 and Sample 2) were used to demonstrate how the Grid is used. Specifically addressed by the MAG are need-to-know issues of concern to strategic managers and systems professionals who are selecting commercial usability evaluation methods for use within their organisation. These need-to-know issues set a minimum standard for what must be addressed by a commercial usability evaluation method.

The Method appraisal grid is a tool that be used by systems acquirers and by systems developers to ensure that the highest quality systems and to demonstrate compliance with the highest usability evaluation practice.

ACKNOWLEDGEMENTS

Thanks are due to colleagues at the Dublin Institute of Technology for their support when preparing of this paper.

REFERENCES

1. Avison, D. and Fitzgerald, G. (1988) *Information systems development.methodologies, techniques and tools*, Blackwell Scientific Publications, Oxford, UK
2. Bevan, N. and Macleod, M. (1994) Usability measurement in context, *Behaviour and Information Technology*, Taylor & Francis Ltd, Basingstoke, UK, Vol(1 & 2)
3. Council Directive (90/270/EEC) (1990) Minimum safety and health requirements for work with display screen equipment", *Official journal of the European Communities*, pp. L 156/14-18
4. Dix, A., Finlay, J., Abowd, G. and Beale, R. (1998) *Human-computer interaction*, Prentice Hall, Hemel Hempstead, UK
5. Fitzpatrick, R. and Higgins, C. (1998). Usable software and its attributes: A synthesis of software quality, European Community law and human-computer interaction, In: *People and Computers XIII. Proceedings of HCI'98 Conference*, Springer, London, UK
6. Hirschheim, R. and Smithson, S. (1988) A critical analysis of information systems evaluation, *Information systems assessment*, Elsevier Science Publishers B.V. North-Holland, pp. 17-37
7. ISO 9000-3 (1997) *International Standard. Quality management and quality assurance standards - part 3:Guidelines for the application of ISO 9001:1994 to the development, supply, installation and maintenance of computer software*, International Organisation for Standardisation, Genève, Switzerland
8. 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
9. ISO/DIS 13407 (1997) *International Standard - Human-centred design processes for interactive systems*, International Organisation for Standardisation, Genève, Switzerland
 10. ISO/IEC 12207 (1995) *International Standard. Information technology - Software life cycle processes*, International Organisation for Standardisation, Genève, Switzerland
 11. Karat, J. (1997) User-centred software evaluation methodologies, Handbook of HCI, 2nd edition, Elsevier Science, p689-704
 12. Kirakowski, J. (1995) "The use of questionnaire methods for usability assessment", *Unpublished collection of readings*, Human Factors Research Group, University College Cork, Ireland
 13. Lim, K.Y. and Long, J.B. (1994) *The MUSE method for usability engineering*, Cambridge University Press, Cambridge, UK
 14. Lin, Han X., Choong, Yee-Yin and Salvendy, Gavriel (1997) A proposed index of usability: a method for comparing the relative usability of different software. *behaviour and Information Technology*, 16(4/5) p267-278
 15. McCall, J.A., Richards, P.K. and Walters, G.F. (1977) *Factors in software quality*, Vols I-III, Rome Aid Defence Centre, Italy
 16. Muller, N., Haslwanter, J.H. and Dayton, T. (1997) Participatory practices in the software lifecycle, Handbook of HCI 2nd edition, Elsevier Science, p255-280
 17. Nielsen, J. (1993) *Usability engineering*, Academic Press Limited, London, UK
 18. Paulk, M. C., Curtis, B., Chrissis, M. B., and Weber, C. V. (1993) Capability Maturity Model for Software, Version 1.1, CMU/SEI-93-TR-24, Software Engineering Institute, USA
 19. Porteous, M., Kirakowski, J. and Corbett (1993) *SUMI users handbook*, Human Factors Research Group, University College, Cork, Ireland
 20. Preece, J., Rogers, Y., Sharpe, H., Benyon, D., Holland, S. and Carey, T. (1994) *Human-computer interaction*, Addison-Wesley, Wokingham, UK
 21. 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
 22. Robson, W. (1997) *Strategic management and information systems:an integrated approach*, Pitman Publishing, London, UK
 23. 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