Towards a Holistic Tool for the Selection and Validation of Usability Method Sets Supporting Human-Centered Design

Holger Fischer[1], Benjamin Strenge[1], and Karsten Nebe[2]

[1] University of Paderborn, C-LAB, Fuerstenallee 11, 33102 Paderborn, Germany
holger.fischer@c-lab.de, best@mail.upb.de
[2] Rhine-Waal University of Applied Sciences, Suedstraße 8, 47475 Kamp-Lintfort, Germany
karsten.nebe@hochschule-rhein-waal.de

Abstract. The establishment of human-centered design within system development processes is still a challenge. Numerous usability methods exist that aim to increase usability and user experience of a system. Nevertheless, the selection of appropriate methods remains to be difficult, as there exist many different factors that have a significant influence on the appropriateness of the methods in their context of use. This paper presents a new concept for the selection of usability methods. It focuses on a) the selection of appropriate usability methods with regard to their applicability in the various stages of system development and b) accounting for interdependencies between multiple methods by balancing them with respect to the usability dimensions effectiveness, efficiency and satisfaction.

Keywords: Human-Centered Design, Usability Engineering, Method Selection, Method Set Validation, ISO/TR 16982.

1 Introduction

Usability has been recognized as an important quality aspect in today's system development industry. However, the integration of usability activities and the establishment of human-centered design (HCD) as part of the development process is still a challenge. Therefore, various approaches have been developed to integrate usability activities in software development, e.g. [1], [2], [3], [4].

Today, many methods exist that aim on an increase of the system's usability using different strategies, e.g. Personas [5], Contextual Inquiry [6], Cognitive Walkthrough [7], etc. One can easily find more than 100 methods by just searching the web. However, a recent study in Germany shows that system development companies rarely use usability methods or only just use the same methods in almost every project due to missing knowledge about alternative methods [8]. Reasons for this lack of knowledge transfer from science to practice might presumably be the vast number of existing methods and lack of suitable support in selecting appropriate ones for a specific project. Some methods are unique but in many cases their divergence is

A. Marcus (Ed.): DUXU/HCII 2013, Part I, LNCS 8012, pp. 252–261, 2013.

minimal and they often make use of an incoherent terminology across the various authors.

The challenge addressed in this paper is to provide tool-based support for the selection of appropriate usability methods, not just by selecting formal and informal criteria (e.g. time constraints, access to test users etc.) but also by considering the applicability in the various stages of system development as defined in ISO/IEC 12207 [9] and interdependencies between multiple methods. The author's approach incorporates a method to validate whether the selected method set is sufficiently balanced with respect to the different dimensions of usability: effectiveness, efficiency and satisfaction.

The paper is structured as following: First, the authors present previous approaches regarding the selection of usability methods and highlight potential for improvements as well as further requirements for a selection concept. Then, the authors report on their proceedings and their method selection approach. Finally, an initial evaluation of the concept and future work is described.

2 Related Work

Several different tools for the selection of usability methods exist. Deliberately excluding static method catalogues, e.g. [10], [11], [12], [13], [14], the following section describes recent dynamic approaches

- 'UsabilityNet' [15],
- 'Usability Planner' [16],
- ISO/TR 16982 [17],
- 'UCD Toolbox' [18],

which differ primarily in terms of the used selection criteria.

2.1 Existing Approaches

Within the 'UsabilityNet' [15] project, 35 methods have been selected based on personal experience of the project partners. These methods are assigned to six system lifecycle steps, i.e. the five steps mentioned in the TRUMP project and a further sixth step of testing and measuring. The methods can be filtered based on three criteria (limited time/resources, no direct access to users, limited skills or expertise of the person executing the method). The selection procedure is therefore quick and easy, but at least one of the three criteria is disputable: There is no reference project for the definition of time and resources although this criterion depends on the complexity of a concrete project. Ten man-days can be very rare for implementing complex accounting software while this is a lot of time implementing a simple website with only six to ten subpages. Actually, 'limited time/resources' is a criterion that should in principle be always affirmed by a project manager because time and resources are always rare and should be minimized in nearly every project.

The ‘Usability Planner’ [16] is a tool, which enables the method selection based on project and organizational constraints. Aspects that have been taken into account are the person’s background (usability expert or software developer), the project stage (e.g. requirements, design, evaluation) and project constraints (e.g. user involvement, task complexity). Methods are filtered by a set of predefined rules and ratings based on the authors’ experiences from practice. This form of subjective rating by a few individuals causes a lack of transparency and may influence the reliability of the selection process.

The ISO/TR 16982 [17] (abbreviated with TR hereafter) includes extensive tables, which rate the appropriateness of several types of usability methods (e.g. ‘document-based methods’, ‘model-based methods’ and ‘creativity methods’) with respect to 18 separate criteria (rated on a scale with five dimensions from ‘recommended’ to ‘not applicable’), such as user involvement, usability expertise, project constraints, lifecycle steps, task characteristics, and properties of the developed product itself. There are also two criteria named ‘Very tight time scale’ and ‘Cost/price control’, which can be disputed analogously to the abovementioned criterion ‘limited time/resources’ from the UsabilityNet method table.

For a given development project, the applicable criteria and the respective ratings for each method category must be determined. However, the TR does not specify how to assess or merge this set of several possibly different ratings for each method category into an overall rating. Due to the usage of categories of usability methods instead of concrete methods, this concept also requires a level of usability expertise, which is not necessarily available to all developers and decision-makers.

Weevers [18] developed an interactive website ‘UCD Toolbox’ on which the selection of methods is primarily based on four main selection criteria: product type (e.g. interface, tangible product), research goal (e.g. ‘Learn about the context of use’), resources (e.g. time, budget, staff) and additional criteria (e.g. equipment, participants number). The authors of this approach focus on concrete methods instead of methods categories. This seems much easier to understand from a usability newcomers’ perspective. Unlike ‘UsabilityNet’ [15], the methods’ complexity is reasonably quantified in man-hours, but selection is done by filtering the list based on a fixed value of 1-5h. This yields methods that may be performed only isolated within that time.

The existing tools’ different selection criteria reflect different approaches to modeling the real world. In order to choose appropriate usability methods, the development project’s characteristics should ideally be captured by the criteria as precisely as possible. This is desirable to adequately rate the individual usability methods. However, due to reality’s complexity it is impossible to determine a finite set of criteria, which cover every aspect. Furthermore, an extensive set of criteria would make the process of method selection very complicated.

2.2 Potential for Improvement and Requirements

The existing tools are useful to select single, decoupled methods for a project. However, the major flaw shared by these approaches is the missing continuity of the methods' usage in terms of a human-centered design process perspective. Usability should not be considered as a one-dimensional construct but rather as a result of many different properties of a system. In practice there is usually not one 'best' method [19], which is sufficient to satisfy all usability goals. Therefore, the outcome of a usability method selection process should strive to provide a set of methods that collectively cover all relevant aspects of a system's usability. It is therefore necessary to provide methods that act in concert by taking the correlations and interrelations of methods into account. Furthermore, the set of methods must actually be executable within the given project's constraints. Obviously, an important aspect is that the selected usability methods match the available project resources.

Apart from this, the knowledge transfer from scientific research into practical usage should explicitly be supported. It must be possible to easily add and consider new and so far unknown methods. The tool should provide enough information about these methods in order to enable usability specialists and developers to perform them with as little effort as possible. Furthermore, it must still be possible to quickly decide which methods to use from an extensive collection of usability methods.

3 Proceedings

The authors have analyzed the existing selection approaches presented previously. The aim was to create an overview of used selection criteria in order to decide about their relevance. Subsequently, the authors developed a concept consisting of several parts. A necessary basis for the selection of usability methods is to establish a collection of methods and adequately describe them. Therefore, a scheme was created for the description of methods based on the specified selection criteria and an initial paper-based collection of fifteen methods has been established. Secondly, the authors developed a concept for a selection tool that is meant to support extensibility of the method collection and the HCD process planning by taking resource requirements in terms of man-days into account. Afterwards, the authors established an algorithm for the validation of the selected method set. The goal was to ensure that the method set takes all dimensions of usability into account. Finally, the authors evaluated the method-rating with usability experts by asking them to select methods out of the initial collection with respect to different given descriptions of development scenarios out of ISO/TR 16982. The output has been compared with that of the simulated selection tool concept using spreadsheet analysis.

As a result, the approach was partially approved and seems to have potential for further research. The detailed proceedings will be described as follows.

3.1 Establishing the Method Collection

The fundamental approach is to enhance the concept of the TR in a way that the appropriateness of extensive collections of concrete usability methods can quickly be rated. Therefore, each concrete method has to be assigned to one of the TR's method categories. The method then inherits the category's ratings with respect to each selection criterion as stated in the TR and may be adjusted whenever research results suggest that the category's 'default' ratings are inaccurate for that method. Because a finite set of selection criteria can never cover all possibly relevant aspects, any additional relevant properties of a method, e.g. applicability constraints, should be provided in textual form. This allows the tool's user to manually take them into account. This is especially important concerning interdependencies with other methods.

Furthermore, the average amount of resources in terms of man-days required for a single execution of the method must be determined. The best way to do so is possibly to have a community-based implementation of the tool to collect as many values from real-life applications as possible and calculate an average. This contributes to replacing the TR's 'very tight time scale' and 'cost/price control', rendering these criteria dispensable.

The concept also incorporates a later validation of method sets to check whether the usability dimensions (effectiveness, efficiency, and satisfaction) are equally well considered (see Section 3.3). The concept therefore claims it is necessary to determine which of these three dimensions are addressed by each method. Such a mapping for some common usability methods can be found in [20].

Fig. 1 (left) shows a wireframe illustration of how a GUI for adding a method to the tool's collection might look like.

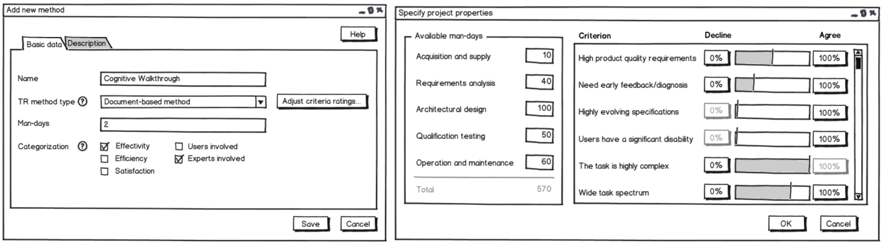

Fig. 1. Wireframes: Adding methods to the collection (left), Entering project characteristics (right)

3.2 Method Selection

To determine appropriate methods for a given development scenario, the project characteristics must be stated in terms of the selection criteria. To resolve the conflict between reality's complexity and usability of the selection tool, the authors adapted the set of criteria specified in ISO/TR 16982 because they are based on expert consensus and are comprehensive enough to enable accurate ratings. Fig. 1 (right) shows a wireframe of a possible GUI for entering this information. As another enhancement to the concept of the TR, the tool's user should not just be able to agree or decline any criterion in a 'binary' way, but state the degree of affirmation, e.g. from a scale of five values. This partially mitigates the issue that for several qualitative criteria of the TR (e.g. 'The task is highly complex') it is sometimes hard to decide whether to apply or to reject. Furthermore, the continuous characteristics of real-world projects can potentially be modeled more precisely this way.

In addition, the available resources (man-days) for each development phase must be entered.

For further calculations, the respective ratings for each criterion must be interpreted as appropriate numeric values (e.g. 'Recommended (++)' → 1; 'Appropriate (+)' → 0.75; 'Neutral' → 0.5; 'Not recommended (-)' → 0).

Based on this mapping, an overall appropriateness rating for each method can be determined by calculating the weighted average of all the ratings with respect to the degree of affirmation to the respective criterion. In the case that any criterion that (partially) applies to the project rates a method as 'Not applicable (NA)', the overall rating for that method must always be the lowest possible value, i.e. zero. This approach enables sorting an extensive collection of usability methods by the overall ratings.

It can then quickly be decided which (of the top-rated) methods to use. Fig. 2 shows an exemplary wireframe of a GUI, which allows the user to state how often each method should be used within each development phase. More than one application of a method within the same phase may for example be necessary when the (iterative) human-centred design process recurs several times within the same step of the system development lifecycle. The sum of man-days required for execution can then automatically be subtracted from the total available man-days. This facilitates the HCD process planning while still leaving room for manual decisions based on properties not directly regarded by the rating algorithm.

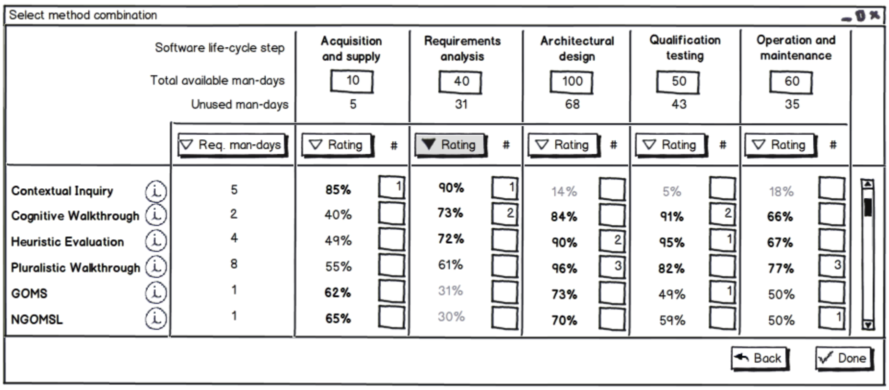

Fig. 2. Wireframe: Method set selection

3.3 Validation of Method Sets

Apart from the calculation of numeric overall ratings for concrete methods and replacing resource-related selection criteria with planning based on man-days, the second major conceptual improvement with respect to the TR is a subsequent validation of the selected method set. Following the above argumentation to illustrate the concept, the usability dimensions (effectiveness, efficiency and satisfaction) are

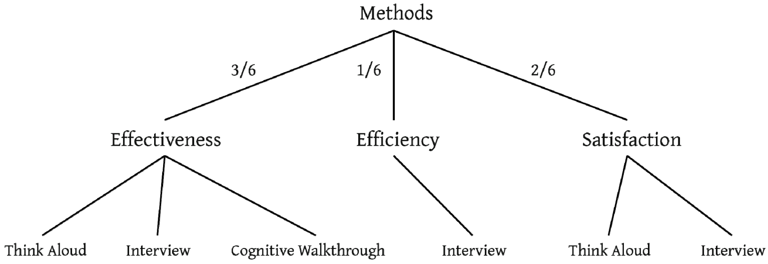

Fig. 3. Method categorization tree

exemplarily regarded as reflecting all aspects of usability as a complex construct. The selected methods can then be interpreted as leafs in a tree of this categorization. Fig. 3 shows an example where the selected method set consists of the methods 'Think Aloud', 'Interview' and 'Cognitive Walkthrough'.

The balance of distribution of the selected methods on the categories (i.e. usability dimensions) can be determined using an inequality metric, e.g. the normalized Shannon entropy

$$H/H_{max} \; = \; -\sum_{k=1}^{n} p_k \cdot \log_n p_k \tag{1}$$

with n = (number of categories). For the example in Fig. 3, this yields a value of

$$H/H_{max} = -(\tfrac{3}{6}\log_3\tfrac{3}{6} + \tfrac{1}{6}\log_3\tfrac{1}{6} + \tfrac{2}{6}\log_3\tfrac{2}{6}) \approx 0,9206 \tag{2}$$

The larger the entropy, the more balanced the distribution in the tree, and the better is the coverage of the different aspects of usability. A value of 1 corresponds to a perfectly equal distribution (best), while a value of 0 means complete agglomeration of all methods within one usability dimension (worst). However, it is difficult to assign a meaningful interpretation to a value $0 < x < 1$. For this reason, a threshold for the entropy must be determined a priori.

When the entropy for a selected method set falls below this threshold, the tool is supposed to inform the user that the selection is not sufficiently balanced. To facilitate corresponding adjustments, the tool should also announce the unattended or accentuated usability dimensions and possibly adequate alternative usability methods.

3.4 Evaluation with Usability Experts

Until now, two features of the method selection concept have been evaluated empirically: The numerical calculation of overall appropriateness ratings for concrete methods and the entropy-based validation algorithm. Thus, an initial collection of 15 representative usability methods was created, containing all of the abovementioned information relevant to assess the appropriateness and validity. A selection based on man-days is dependent on properties of real world projects, which can hardly be simulated for evaluation purposes. For this reason, that conceptual feature was excluded in the evaluation.

As a reference source for evaluation, five real-world development scenarios presented in ISO/TR 16982 Annex B (‘Examples of in situ applications’) have been used. Each consists of a short description of the project goals and constraints in natural language, and the corresponding set of TR selection criteria. Due to the fact that these given criteria must be regarded as ‘fully agreed upon’ with respect to the scenario (fully declining all other criteria), the concept of allowing the user to partially affirm a criterion did not come into play in this first evaluation.

The descriptions have been exempted from any specific hints related to the usability methods that had actually been used in the real projects. Then, the descriptions have been given to usability professionals unfamiliar with the selection tool concept. The experts were each independently asked to name the methods they would apply in each scenario.

Sometimes an expert stated to be unsure which of some alternative methods to use when several methods were recognized as appropriate. The preferred methods were then assigned the value ‘1’, alternative methods were weighted with ‘0.75’ (according to the abovementioned value for criteria with an ‘Appropriate (+)’ rating), and unused methods were given the value ‘0’. As can be seen in Fig. 4, there was a strong tendency of methods with a higher calculated overall rating also being chosen more likely by the usability experts.

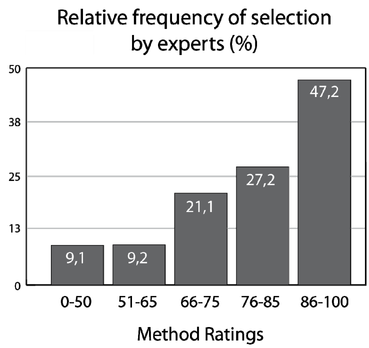

Fig. 4. Evaluation results

Surprisingly, there were also a few cases where methods have been rated ‘0’ overall by the simulated tool, but still chosen by the professionals. This was probably due to the fact that the experts sometimes came to an intuitive understanding of the project characteristics based on the textual descriptions that differed significantly from the facts covered by the selection criteria.

Although man-days-based planning was not explicitly included in the evaluation, this concept interestingly got indirect support from the fact that all experts stated that the time frames specified in the scenario descriptions were in most cases highly important for their choices. Actually, the professionals mostly started with an analysis of how much time is available and which methods could possibly be used under this condition.

4 Conclusion and Outlook

In this paper, the authors presented a concept for the establishment of a community-based usability method collection as well as for the selection of an adequate method set for a user-specified project. A first evaluation of the basic method-rating concept was done using spreadsheet calculation and reviews from usability experts.

The author's analysis of existing approaches to the systematic selection of usability methods showed that usually a narrow angle is taken. In these approaches single methods are selected in isolation from the larger context of planning the development process is taken. Although usually several different methods must be used within one project, their interdependencies are not taken into account. To improve the knowledge transfer from science to practice, a selection tool's method collection should be extensible in an unproblematic and systematic way. Methods should be described in a way that HCI specialists as well as developers can easily understand and learn how to perform each method (suitability for learning).

The current status of our work on these issues is primarily a theoretical concept for method selections based on and extending the ISO/TR 16982. The approach allows calculating numeric overall ratings for the appropriateness of concrete usability methods based on a flexible and robust specification of project characteristics and other criteria. When a computer automatically does this calculation, it enables the user to easily assess extensive collections of usability methods. To facilitate the HCD process planning, the concept envisages quantifying available resources and the methods' requirements in terms of man-days.

In order to fortify the holistic consideration of all aspects of usability during a development project, the concept also provides an algorithm to validate selected method sets regarding their balance. The evaluation results suggest that the calculated overall ratings of concrete methods coincide with judgment from usability professionals. However, until now the low number of samples prohibits a definite conclusion.

Future work will contain further evaluations and an implementation of the concept in form of a conventional software tool or rich Internet application. Mechanisms for feedback from a community will be integrated and used to refine information in the methods repository. Such a loopback is expected to be especially beneficial to determine the man-days required for method executions but also for the categorization of methods and the adjustment of criteria ratings.

Finally, apart from practical applications, the concept presented in this paper could lead to further general theoretical discussions regarding method selection approaches, e.g. the ongoing revision of ISO/TR 16982 to ISO 9241-230.

References

1. Lallemand, C.: Toward a Closer Integration of Usability in Software Development: A Study of Usability Inputs in a Model-Driven Engineering Process. In: Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems EICS, pp. 299–302. ACM, New York (2011)
2. Fischer, H., Nebe, K., Klompmaker, F.: A Holistic Model for Integrating Usability Engineering and Software Engineering Enriched with Marketing Activities. In: Kurosu, M. (ed.) Human-Centered Design, HCII 2011. LNCS, vol. 6776, pp. 28–37. Springer, Heidelberg (2011)

3. Nebe, K., Zimmermann, D., Paelke, V.: Integrating Software Engineering and Usability Engineering. In: Pinder, S. (ed.) Advances in Human-Computer Interaction, pp. 332–350. InTech (2008)
4. Seffah, A., Desmarais, M.C., Metzker, E.: HCI, Usability and Software Engineering Integration: Present and Future. In: Seffah, A., Gulliksen, J., Desmarais, M.C. (eds.) Human-Centered Software Engineering – Integrating Usability in the Development Process, pp. 37–57. Springer, Heidelberg (2005)
5. Adlin, T., Pruitt, J.: The Essential Persona Lifecycle – Your Guide to 'Building and Using' Personas. Morgan Kaufmann, San Francisco (2010)
6. Holtzblatt, K., Wendell, J.B., Wood, S.: Rapid Contextual Design – A How-to Guide to Key Techniques for User-Centered Design. Morgan Kaufmann, San Francisco (2004)
7. Benyon, D.: Designing Interactive Systems: A comprehensive guide to HCI and interaction design, 2nd edn. Pearson Education Ltd., Harlow (2010)
8. Woywode, M., Mädche, A., Wallach, D., Plach, M.: Usability of Software Applications as a Competitive Factor for Small and Medium-Sized Enterprises. German Federal Ministry of Economics and Technology, BMWi (2012) (in German)
9. ISO/IEC 12207: Systems and software engineering – Software life cycle processes (2008)
10. Usability Body of Knowledge, `http://www.usabilitybok.org/methods`
11. Generic Work Process, `http://project.cmd.hro.nl/cmi/hci/toolkit/index2.php`
12. Usability Gov, `http://www.usability.gov/methods`
13. The Methods Lab, `http://www.education.edean.org/pdf/Tool039.pdf`
14. Ideo Method Cards, `http://www.ideo.com/work/method-cards`
15. Bevan, N.: UsabilityNet Methods for User Centred Design. In: Jacko, J.A., Stephanidis, C. (eds.) Human-Computer Interaction: Theory and Practice (Part 1), HCII 2003, vol. 1, pp. 434–438. Lawrence Erlbaum Associates, Mahwah (2003)
16. Ferre, X., Bevan, N.: Usability Planner: A Tool to Support the Process of Selecting Usability Methods. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 652–655. Springer, Heidelberg (2011)
17. ISO/TR 16982: Ergonomics of human-system interaction – Usability methods supporting human-centred design (2002)
18. Weevers, T.: Method selection tool for User Centred Product Development. Master Thesis, Delft University of Technology (2011)
19. Wixon, D.: Evaluating Usability Methods: Why the Current Literature Fails the Practitioner. Interactions 10(4), 28–34 (2003)
20. Freymann, M.: Classification of User-Centered Evaluation Methods within the User-Centered Design Process. Diploma Thesis, University of Paderborn (2007) (in German)