



HAL
open science

Predictive Usability Evaluation: Aligning HCI and Software Engineering Practices

Káthia Marçal de Oliveira, Sophie Lepreux, Christophe Kolski, Ahmed Seffah

► **To cite this version:**

Káthia Marçal de Oliveira, Sophie Lepreux, Christophe Kolski, Ahmed Seffah. Predictive Usability Evaluation: Aligning HCI and Software Engineering Practices. IHM'14, 26e conférence francophone sur l'Interaction Homme-Machine, Oct 2014, Villeneuve D'Ascq, France. ACM, IHM'14, pp.177-182, 2014. hal-01090450

HAL Id: hal-01090450

<https://hal.science/hal-01090450v1>

Submitted on 3 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predictive Usability Evaluation: Aligning HCI and Software Engineering Practices

Káthia Marçal de Oliveira

UVHC, LAMIH CNRS UMR 8201
Le Mont Houy cedex 9
59313 Valenciennes, France
kathia.oliveira@univ-
valenciennes.fr

Sophie Lepreux

UVHC, LAMIH CNRS UMR 8201
Le Mont Houy cedex 9
59313 Valenciennes, France
sophie.lepreux@univ-
valenciennes.fr

Christophe Kolski

UVHC, LAMIH CNRS UMR 8201
Le Mont Houy cedex 9
59313 Valenciennes, France
christophe.kolski@univ-
valenciennes.fr

Ahmed Seffah

UVHC, LAMIH CNRS UMR 8201
59313 Valenciennes, France
ahmed.seffah @univ-
valenciennes.fr
Lappeenranta University of
Technology (LUT)
P.O Box 20 FI-53851
Lappeenranta, Finland
ahmed.seffah@lut.fr

Abstract

Can we - software developers, usability experts, user interface designers - predict usability from the early user interface (UI) design artifacts and models? Can we define predictive measures to evaluate usability without a concrete UI? These questions seemed natural for us since UI modeling (task, user, concepts, etc.) is being largely explored in recent years for the automatic generation of final UI. To answer those questions we propose a model-based predictive usability evaluation approach that uses a set of usability measures. These measures are the essence of a framework we are developing for usability prediction. Initial empirical studies were performed to support this approach. This paper presents the fundamental basis on top of which we have developed this approach.

Keywords

Usability, measures, evaluation

Keywords – ACM classification

H.5.2. Information interfaces and presentation (e.g., HCI): User Interface – Evaluation/Methodology.

Introduction

During the last two decades, usability evaluation has been intensively treated by Human-Computer Interaction (HCI) researchers and usability professionals. A large number of usability evaluation

© ACM, 2014. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 26^{ième} conférence francophone sur l'Interaction Homme-Machine, 2014.

<http://dx.doi.org/10.1145/2670444.2670467>

methods and models have been proposed. An exhaustive account of these methods can be found in [4]. Several models and standards ([2], [3]) have been established to measure usability as a quality construct. Mainly these models break down usability into measurable attributes. They also provide a list of measures that are generally in conjunction with usability evaluation methods to qualitatively and quantitatively estimate usability attributes. Building on or contributing to these standards and usability evaluation methods, several European projects suggested specific frameworks to measure usability; representative examples include SUMI (Software Usability Measurement Inventory) and MUSiC (Metrics for Usability Standards in Computing) developed by the Human Factors Research Group, University College, Cork. Trying to integrate different models in a same structure from a literature review, Seffah et al. [11] proposed a consolidated model, named QUIM (Quality in Use Integrated Measurement). QUIM is a unifying model that assesses usability by using 12 factors, 29 criteria and more than 120 measures. Another significant research is the Single Usability Measure suggested to standardize usability measure into a single global score [9]. Moreover, guidelines and ergonomic criteria for usability evaluation were also proposed (e.g. [10]).

However, most of those projects do not address explicitly usability prediction from early UI design models. This is because the calculation of the proposed measures in these projects requires a fully functional prototype. Among the few initiatives, one can mention, the European action COST294 that highlighted the need for theoretical frameworks defining the nature of interaction quality. This includes techniques and

measures for the evaluation of the interaction quality during design. Another initiative is the proposition of Panach et al. [7] to represent usability features related to functionality abstractly in a conceptual model, and then to carry out automatically their implementation. To that end, they proposed a method to define the functional usability features (e.g. feedback support, user input error prevention, go back) in a model-driven approach using what they call conceptual primitives. This method includes the identification of usability mechanisms, changes in conceptual models and in the compilers. Measures are also proposed for early usability evaluation considering these conceptual primitives. To interpret the result of the measures, they assign categorical values (e.g. very good, good, bad) for range of numerical values based on guidelines in the literature. However, the classification of measure results into categories still remains intuitive and is not correlated with the end user view. Considering this scenery, we plan to work on a definition of a framework for predictive usability evaluation from early design models based on measures that can be interpreted using the perception of the users about the usability of final user interfaces. No extra effort is required in the UI design to allow the quality evaluation.

Next sections present our approach to perform predictive usability evaluation, a preliminary investigation on this direction and some final remarks.

Definition of an Approach for Predictive Usability Evaluation

The context we envision is that UI designers can, during UI modeling and early design phases, apply the specific measures related to that model and have some indication of the level of usability of the final UI using

pre-defined correlations. The new UI being developed and evaluated can later also have its measures integrated in the historical base of evaluations, thus improving the correlations.

To that end, we have planned to define a framework that consists of a set of measures applicable to different UI design models from a high abstraction level to the UI implementation [6]. The prediction of usability and the analysis of measures results are based on the possible correlations between predictive usability measures (on the design artifacts) and the results of usability tests performed by users (on the final UI). The correlations will be defined with several empirical studies.

Figure 1 portrays the proposed predictive usability prediction approach as well as the way it can supplement inspection-based evaluations. Predictive usability evaluations are performed on various design models using usability predictive measures.

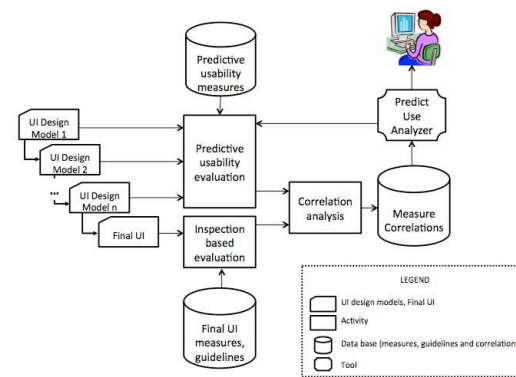


Figure 1. Predictive Usability Evaluation Approach

The predictive usability evaluation uses a database including the correlations gathered from previous empirical experiments. Each experiment defines and refines continuously the correlations between the results of usability measures collected from early UI design models and the results of inspection-based evaluations made on the final UI. A predictive usability analyzer tool (PredictUse Analyzer) will support the measures collection and calculation as well as the usability prediction that uses the database.

A preliminary investigation

We had surveyed different quality models that propose a list of measures. Our starting point was the QUIM model [11] since it is defined considering several quality models and it organizes a set of usability measures. While investigating those measures, we noted that most of the measures were applicable only to final UI since the main goal of QUIM is the definition of quality in use that means, while using the UI. We decided therefore to analyze also the internal measures of usability (applied to a non-executable software product) from ISO9126. The ISO9126 external measures has already been included in QUIM. Finally, we also analyzed the ergonomic criteria from the most cited usability model, defined by [10].

Our procedure was first to read the measure definition (description, formula and scales when available). Next, considering the features of each design model, we identified if that measure could be applied to that design model even that the formula customization would be needed. Finally, we analyzed the measures against the three UI design models proposed in the CAMELEON framework [1]:

- Task model - that describes the various users' tasks to be carried out with a UI. Concur Task Tree (CTT) [8] is the most used notation to define this model;
- The abstract User Interface - that defines the rendering of the task model independently from any style of interaction (e.g. graphical interaction, vocal interaction, speech synthesis and recognition) as well as the capability and constraints of the platform;
- The concrete user interface – that concretizes an abstract UI into an interactor-dependent expression for a given context of use;
- Final UI – that is the operational user interface running in a specific platform.

To do the analysis, we took into account the features of each one of the UI models. For instance, the task model defined using CTT, is composed basically of a hierarchy of different kind of tasks (abstract, interaction, user and application tasks) and temporal relationships among tasks (interleaving, synchronization, enabling, enabling with information passing, choice, deactivation, iteration and optional).

Table 1 presents some measures extracted from QUIM. We discovered that some measures require the execution of the UI (e.g. temporal efficiency), therefore, it will be applied to the final UI only. Others are related to the UI design itself (e.g. local density), are thus related to the concrete UI and some of them to abstract UI. Finally, some measures are related to the information to be presented in the UI (e.g. interface shallowness), thus they can be applied in all UI design models from the highest level of abstraction (the task model). After analyzing measurable criteria

(considering QUIM measures, SQUARE standard, and Scapin and Bastien criteria [10]), we started the customization of them to evaluate each one of the models.

Table 1. Some examples of the analysis of measures

Measures	Task model	Abstract UI	Concrete UI	Final UI
Function understandability	✓		✓	✓
Human Efficiency	✓		✓	✓
Interface Appearance Customizability			✓	✓
Interface shallowness	✓	✓	✓	✓
Layout Appropriateness		✓	✓	✓
Local Density		✓	✓	✓
Longest Depth	✓	✓	✓	✓

Table 2 presents the analyses of some measures applicable to task model. We note that we have qualitative and quantitative measures. The qualitative measures are collected by inspections with UI designers. Those inspections are cognitive walkthrough based [5], since the assessment is done by experts who should wonder the potential perception of the user in the use of UI implemented based on that task model. The quantitative measures can be calculated from the data obtained from task models tools (e.g. CTTE).

It is important to highlight that each one of those measures is not complete, that means, each measure is defined to measure what can be measured in the task model considering only the elements of this model. However, in other models and in the final UI other aspects of the same measure should be evaluated. For example, the measure defined for information density considers only the aspect number of input/output in the same task since this task will probably be designed in

the same panel/window (if we consider graphical UI). However, when evaluating this measure in an abstract UI or in a concrete UI we can look for the organization of those input/output related to the whole space of the panel/window.

Table 2. Examples of measure for task model

Measure	Measurement
Function understandability	Subjective evaluation with five point ordinal scale (1-5)
Longest Depth	Computes the number of nodes in the CTT hierarchy from the root till the primitive task
Visual Coherence	$100 - \text{Number of primitive task that are not closely related in the same high level task} / \text{number of primitive tasks defined} * 100$
Minimal Actions	Compute the minimum path of access to primitive task
Input validity data	$(\text{Number of sequence of "Enabling with information passing"} + \text{number of tasks explicitly defined to validate the data}) / \text{Number of tasks is decomposed in at least one primitive task}$

We also highlight that some transversal features cannot be evaluated in a task model since they are usually not modeled. These transversal features include: help, interface customization aspects, specific aspects of navigability such as home, exit, etc.

Finally, we note that our idea is not to define an interpretation of each measure. This will be done by comparing the result of evaluation for a specific project with other results plotted in a graph of correlation points. This graph correlates measures of models with the evaluation of final UI.

To allow the prediction evaluation we need to compare these results with historical data of previous evaluations where correlations were discovered between the evaluation of the task models and the final UI. The definition of a database of correlation takes time since we need several UI designs of several projects, the definition of the UI based on those models and the evaluation from UI design models and final UI.

However, to evaluate the feasibility of our idea, we performed a case study with students of the second year of the two-year graduate course on computer science at the University of Valenciennes. Although we had little data (collected from 14 projects done by students during the course) for an adequate predictive evaluation, we would like to show an example of the proposed approach. Figure 2 presents a graph that shows the evaluation of the task model (x axis) for the measure input validity data and the evaluation of the same measure in a final UI (y axis) (we used a scale from 1 (input never validated) to 5 (input always validated)). It shows that when the task model has a good evaluation the final UI tends to have a better evaluation too. We obtained a value of Spearman correlation equals to 0.46, which is not very good, but that we considered acceptable since we do not have a lot of points (14 projects) and the subjects (i.e., the students) had very little experience in task modeling and programming.

This preliminary study showed us that there a lot of work need to be perform to effectively validate this approach since we must have UI design models and the final UI developed based on those models.

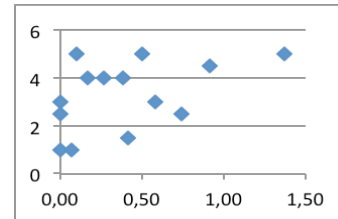


Figure 2. Evaluation of UI design model (x axis) x Final UI for input validity data (y axis)

Conclusion and Future Work

This paper presents the first results of a long-term research that aims to define a predictive usability

References

- [1] Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J. A Unifying Reference Framework for Multi-Target User Interfaces. *Interacting with Computers*, v. 15, n. 3 (Jun 2003), 289–308.
- [2] ISO 9241-11. Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on Usability (1998).
- [3] ISO/IEC 25023. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality (Aug 2011).
- [4] Ivory, M. Y., Hearst, M. A. The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, v. 33, n. 4 (Dec 2001), 470-516.
- [5] Mahatody, T., Sagar, M., Kolski, C. State of the Art on the Cognitive Walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, 26 (8)(2010), 741-785.
- [6] Oliveira, K. New Research Challenges for User Interface Quality Evaluation. *Actes d'IHM 2010*, (Sep 2010), 145-148.

framework to be used in UI design. The main advantage of this approach is that the interpretation of the quality evaluation of the UI design models by measures is supported by previous evaluations of the end user in similar projects. The main limitation of this approach is the need of previous correlation analysis to define the prediction of usability. To have good correlations, a large number of projects is needed to evaluate the design models and the final UI. We are now working on the definition of the database of correlations with controlled projects in the computer sciences courses. Future works include the complete definition of the framework and its application in large projects.

- [7] Panach, J. I., Juzgado, N. J., Pastor, O. Including functional usability features in a model-driven development method. *Computer Science and Information System Journal*, v.10, n.3 (2013), 999-1024.
- [8] Paterno, F., Mancini, C., Meniconi, S. ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models, *INTERACT '97* (Jul 1997), 362-369.
- [9] Sauro, J., Kindlund E. A method to standardize usability metrics into a single score. *CHI'2005* (2005), 401-409.
- [10] Scapin, D.L. and Bastien, J.M.C. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, v. 16, n.4 (1997), 220-231.
- [11] Seffah, A., Donyaee, M., Kline, R. Padda, H. Usability measurement and measures: A consolidated model, *Software Quality Journal*, v. 14 (2006), 159-178.