

Evaluating Quality and Usability of the User Interface: A Practical Study on Comparing Methods with and without Users

Caio Marcio Silva^{1,*}, Vanessa Macedo¹, Rafaela Lemos²,
and Maria Lúcia L.R. Okimoto²

¹ Instituto de Pesquisa para o Desenvolvimento LACTEC and
Universidade Federal do Paraná UFPR, Curitiba, Brasil

² Universidade Federal do Paraná UFPR, Curitiba, Brasil
{vanessa.macedo, caio.silva}@lactec.org.br,
rafaelalemos.dgn@gmail.com, lucia.demec@ufpr.br

Abstract. This paper aimed to verify the suitability of the methods “Rehearsal Technique” with users and “Heuristic Evaluation” in the evaluation process of the usability of appliance’s control panel, during the product development process. Both methods were applied with the same goal, identify usability issues with emphasis on information design, and with the same product, a functional prototype of a washing and drying machine’s control panel. The heuristic evaluation was applied by three analysts, and the rehearsal technique, applied with ten participants. Each of the methods resulted on usability issues, which were evaluated by criticality and related to design fields. It was concluded that the heuristic evaluation allowed the verification of broader usability problems, early detecting about 85% of the issues identified on the rehearsal technique. However, the rehearsal technique enabled the identification of user feedback, as well as actual behaviors, and confirming some usability issues. Therefore, it is indicted the consideration of heuristic evaluation as a helpful strategy on evaluation of usability, not only as a preliminary evaluation for the usability test.

Keywords: user experience, home appliance, usability assessment.

1 Introduction

Assessing products usability is a subject discussed in the literature and also in the context of research and development in some companies. In this sense, Maguire (2001) [11] points out some benefits of a good usability system, such as increased productivity, reduced number of errors during the interaction, the decrease of necessary support and training, as well as improvement of accepting the product. As a consequence of this, there is an improvement of the way in which the company is seen by the user. Methods, with emphasis on usability, are: discussed by authors (Karwowski et. al., 2011; Dumas & Salzman, 2006) [10], [4], presented in "manuals"

* Corresponding author.

(Stanton et. al., 2005) [15]; and indicated by standards (ISO 16982:2002) [8]. However, there is a constant concern to identify low-cost methods, fast to be applied and that result in valid data [12]. In this direction, this research intends to help professionals in the field of usability and user experience to select methods for evaluation of products control panels, from the comparative reporting involving two different evaluation methods: one of them with no user involvement (Heuristic Evaluation) and the other with users involvement (Interaction Rehearsal Technique). The object of study was a graphic user interface of a washing and dryer machine.

It aimed to make a comparison between the results of these two methods. Thus, both methods were applied with the same interface and the same purpose: to identify usability issues with emphasis on information design. The results were compared, verifying the appropriateness of the methods for the following evaluation criteria: performance time, number of usability problems identified; criticality of the raised usability problems, and number of usability problems for the information design.

2 Usability Evaluation

Regarding methods with emphasis on usability and related topics, Dumas & Salzman (2006) [4] present methods for evaluating products divided into four categories: (i) inspection methods (heuristic evaluation and cognitive walkthrough), (ii) usability testing, (iii) self-report methods (questionnaires, interviews and focus groups), and (iv) methods of observation (ethnography). Maguire (2001) user-centered design, setting as methods for evaluation: participatory assessment; dynamic assessment; heuristic evaluation; controlled testing with users; satisfaction questionnaires; assessment of mental loading; analysis of critical incidents; and post-experience interviews.

For choosing a usability method, many factors must be taken into account. Stanton & Young (1999) [14] suggest that it should be considered the project developmental stage, available resources and time, in addition to the analysts' skills, access to end users and the necessity of data for the project. Annett & Stanton (2000) point out issues that must be raised for the selection of the usability method, such as: "how deep should the analysis be?"; "what existing tools to provide support for the use of the method?"; "how reliable and valid is the method?", among other questions.

Also in this context, Roto et. al. (2009) [12] conducted a survey with professionals in UX. It was found that professionals who came from industry show more concern about the resources required for implementation of the method than professionals from academia, as well as the speed and ease of application. They also report that qualitative methods are preferred when the product is in the early stages of the development process, seeking to present constructive data to the design team. Moreover, Tullis & Albert (2008) [16] emphasize the need of presenting numerical data to customers at the time of the usability report, making quantitative data a necessity. In this context, it is essential that the usability and UX professional is aware of costs, time and recruitment for applying the methods, as well as the method speed to detect key points of usability problems.

From the literature review, there is a high number of usability methods directed and/or applied to the human computer interaction [6], not being identified a method developed specifically for evaluating the usability of home appliances. Thus, this research was based on general methods, which can be applied in different contexts for different types of products (systems and product with three-dimensional interface). Given the research of Roto et. al. (2009) [12], which highlights usability testing, and of Dumas & Salzman (2006) [4], with emphasis on the evaluation by a specialist, there is the motivation to check the positive aspects and caveats of these two methods.

However, there are many variables in usability testing highlighted by Tullis & Albert (2008) [16], such as: number of observers, systematization of data tabulation (synchronous, asynchronous, both), application of thinkaloud, moderator interventions, support systems for use (instruction manuals, quick guide, instructional videos), use of specialized software for data tabulation, insertion of pre and post interaction questionnaires, among others. Faced with these choices which influence costs and resources for the implementation, it was decided to choose, among the possibilities for usability test, the Interaction Rehearsal Technique. This procedure was selected for having less statistical, technical accuracy (for example: materials, control of variables, number of involved analysts), as well as for allowing an evaluation from the potential users interaction with the product.

Just like Usability Testing has some variants, the assessment by expert also presents alternatives; either the number of experts who will carry out the activity, the type of material that will be used (heuristic, checklist) and the way the activity will be performed (cognitive walkthrough or traditional scan). Still within the same category of review by an expert, there are other options, such as the types of used heuristics, golden rules, principles of dialogue and ergonomic criteria.

The heuristics proposed by Nielsen (1994) suggest aspects to be followed by the developed interfaces. They are: visibility of system status; match between system and the real world; user control and freedom; consistency and standards; help users recognize, diagnose, and recover from errors; error prevention; recognition rather than recall; flexibility and efficiency of use; aesthetic and minimalist design; and help and documentation. The golden rules proposed by Shneiderman & Plaisant (2004) [13], are the following: strive for consistency, enable frequent users to use shortcuts, offer informative feedback, design dialog to yield closure, offer simple error handling, permit easy reversal of actions, support internal locus of control and reduce short-term memory load. There are also the dialogue principles proposed by ISO 9241-11 (1998) [8], which are: suitability for the task, self-descriptiveness, controllability, conformity with user expectations, error tolerance, suitability for Individualization and suitability for learning.

In addition to these authors, there are others who contemplate aspects for evaluation of interfaces from the inspection by experts, among them, Bastien & Scapin (1993) [2]. The authors argue that the interfaces should be developed from some ergonomic criteria which can subsequently be elements for evaluation, such as: conduction; workload; explicit control; adaptability; errors management; uniformity/consistency; and meaning of codes denominations, as well as compatibility. Given this literature review, it was created a table summarizing the topics discussed here and the principal authors.

Table 1. Synthesis of authors raised for the literature review

Area	Authors
UX / Usability Methods	Karwowski (2011); Dunas and Salzan (2006); Tullis and Albert (2008); Roto et. al. (2009)
DCU / Human Factors Methods	Stanton et. al. (2005); ISO 16982:2002; Maguire (2001); Stanton and Young (1999); Annett and Stanton (2000).
Heuristics / Recommendations	Nielsen (1994); Shneiderman and Plaisant (2004); ISO 9241-10:1998; Bastien and Scapin (1993);

3 Method

This study was conducted in four stages. The first step one consisted in the literature review. In step two, the heuristic evaluation by three usability analysts was performed. In step three, the interaction rehearsal technique was performed. Finally, in step four the analysis of results and report writing was done.

In order to provide ways of direct comparison, it was selected the same artifact to be evaluated with heuristic and from the interaction rehearsal technique; it is a control panel of a washing and dryer machine. It aimed to carry out such assessments with the artifact in development, simulating a formative evaluation. For this, it was used an Adobe "flash" file, simulating smaller analysis cycles of the control panel, from tests with functional prototype. However, even without operating the machine. This kind of prototype, in the context of the development of graphical interfaces for home appliances, is one of the most faithful ones, when compared to the prototypes on paper and interactive guides. In it, audible and visual feedback can be more faithfully reproduced.

3.1 Heuristic Evaluation

For the heuristic evaluation, three professionals of usability with products were recruited, with different levels of training. It was then requested that these professionals made individual evaluations, based on Nielsen's heuristics (1994) and on the ergonomic criteria of Bastien & Scapin (1993) [2], completing a worksheet, exemplified in Table 1, and they were free to choose how to lead the heuristic evaluation. This evaluation guides were the present items in the columns of the available spreadsheet: criticality, which is how critical the problem is and needs to be priority to be modified; the description of the problem, pointing the problem to the developer; the Heuristic or Ergonomic Criterion infringed, who moved the problem

description, the area, that directs which development team will be responsible for a solution/improvement for the usability problem, and the impact, which indicates what is the size of usability negatively influenced by this usability problem.

Table 2. Example of the usability problems spreadsheet

Criticality	Problem description	Violated Ergonomic Criterium or Heuristic	Area	Impact
Problem criticality (high/médium/low).	Succint description of identified problem.	Indication of the heuristic violated by the problem.	Relation between the usability problem and one of the three areas: graph, information and interaction.	Delimitation of what dimension(s) of usability can be impacted (satisfaction, efficiency and effectiveness).

3.2 Interaction Rehearsal Technique

After the heuristic evaluation, the interaction rehearsal technique was performed, composed of ten sessions with voluntary users. The participants were chosen according to schooling, so that 50% of the participants had a college degree and 50% had incomplete college degree. Among these, 60% were male and the rest (40%) female. The ages of the participants were concentrated in the range of 21 to 30 years, except for one participant who was in the range of 41 to 50 years. The graduation area of the participants was varied, as examples: engineering, college student, psychology and design.

Each session lasted on average 40min. The interaction rehearsal technique, although with no statistical rigor as data capture and sample selection, aimed to investigate the relationship between users and the artifact under study and, from this, identify usability problems related not only to use but also regarding the design of interfaces.

In the first step, the demographic questionnaire aimed to capture data as education, gender, occupation and age group, while the inspection of familiarity sought to ascertain whether the user had knowledge about the use of this category products. Through this tool, it was investigated if the user had ever had contact with washing and dryer machines, if he/she has such a product in his/her home and what is the frequency of use. After questioning "pre interaction" it started the Step 2 - Rehearsal Technique. At first, the user was placed in front of the product and the moderator, in turn, encouraged the participant to verbalize the meaning of each function of the graphic user interface and its mode of use, even with no physical contact with the interface. Still in this step without tactile interaction, two tasks were also presented to the user. After the utterance of these tasks, the participant was encouraged to submit their mental model from the verbalization of the steps taken to complete them.

The rehearsal was pursued by tactile interaction, composed by presenting four scenarios, addressing tasks with distinct features like: use of "hypoallergenic" function; use the function "delay start"; using a reduced amount of clothing; and use the function "drying", without washing. Finally, it was presented a translated and adapted version of the satisfaction questionnaire after use QUIS (Questionnaire for User Interface Satisfaction).

All the ten sessions were observed by a usability analyst, who performed a synchronous tab. After the session, the data were reviewed and, after these ten sessions, the data and report were sent to two other analysts, responsible for synthesis usability problems. The identified problems were organized into three groups: initial exploration, tasks mental model and tasks, as shown in Table 2. Each behavior obtained a frequency (sum of the number of participants who performed the particular behavior), and from this number, were generated usability issues for those behaviors often equal to or greater than three.

Table 3. Example of spreadsheet of recording behaviors in Interaction Rehearsal Technique

Behavior	P1	P2	...	P10	Frequency	Usability Problem
Initial Exploration						
Mental Model						
Tasks						
Comments						

The usability problem frequency was also related to criticality: usability problems with frequency less than or equal to four were considered low criticality; usability problems often five and six were considered average criticality; and usability problems often more than or equal to seven were classified as high criticality.

4 Results

4.1 Results of the Heuristic Evaluation

As it was mentioned in the method, the Heuristic Evaluation was performed by three usability analysts; here named "A1", "A2" and "A3". The analyst "A1" identified fourteen usability problems, being five of average criticality and nine of low criticality. The analyst "A2" also found fourteen usability problems, of which eight were low criticality and six average criticality. Finally, the analyst "A3" appointed eleven usability problems, among them, five of average criticality and six of low criticality. These data are presented in Table 1.

Table 4. Number of usability problems identified by the three analysts in the Heuristic Evaluation

	A1	A2	A3
Low criticality	9	8	6
Average criticality	5	6	5
High criticality	0	0	0
Total of usability problems	14	14	11

The usability problems verified by the three analysts were grouped into a spreadsheet for a total twenty-five distinct usability problems. Therefore, the incidence of the same usability problem was observed in the evaluation of more than one analyst. Four usability problems of low criticality and two of average criticality were also detected by two analysts. Regarding the agreement of the three analysts in usability issues, a total of four issues were observed, two of which were classified as average criticality and two as low criticality. Divergence was seen in the classification of criticality in just two usability issues, prevailing the classification by the analyst(s) with most experience in this area. Regarding the runtime and analysis of heuristic evaluation, each analyst reported having developed the heuristic evaluation in part time work (four hours) and it was computed a bout of work for data analysis, accounting for the entire eight working hours, requiring three analysts and a final evaluator.

Among the problems that concordance of two or more analysts (total of ten) were found, it appears that the most violated heuristics were: "Communication", with four incidences of problems with compliance, "Conduction and Control" and "User Freedom", both with two incidences. Among the three areas of categorization of usability problems identified in the heuristic evaluation, information design was categorized in all ten problems, interaction design was related to six problems and, finally, graphic design displayed in four problems. Regarding impacted dimensions, efficiency was pointed at eight problems, followed by satisfaction, pointed at four usability issues, and effectiveness, in turn, occurring in three usability problems: as presented in Table 4.

Table 5. Violated heuristics, related areas and dimensions impacted on issues of agreement among analysts

Violated heuristic	Related area	Impacted dimension
Communication (4)	Information (10)	Effectiveness (3)
Workload (1)	Interaction (6)	Efficiency (8)
Conduction (2)	Graph (4)	Satisfaction (4)
Compatibility of system – real world (1)		
Control and freedom for the user (2)		
Consistency and standards (1)		

4.2 Results of Interaction Rehearsal Technique

With respect to technological familiarity of the ten participants, seven have a washing machine in their homes, two participants use collective laundries and one participant does not use this type of product. It was found that 7 participants use the product from one to four times a week, while 3 participants use the washing machine rarely. These three participants, one had never used a washing machine or clothes dryer.

The interaction rehearsal technique was tabbed from the verification of users' behaviors, both errors as hits during the execution of the task presented. A matrix has been developed for the recognition of functions of the graphic user interface, related to the exploration stage with no tactile interaction. For this phase, it was observed that the functions of lower scores were: "lock panel" (4/10 hits); "rinse" (5/10 hits); "spin" (5/10 hits) and "unlock door" (5/10 hits).

Forty-four (44) behaviors were identified in the ten sessions of the interaction rehearsal. The behaviors which obtained frequency equal to or greater than three were transcribed into a format of usability problems. This procedure resulted in a total of fourteen (14) different usability problems. It was noted also that two distinct behaviors can result in a same usability problem.

With respect to criticality of the issues raised in the interaction rehearsal, it was verified four usability problems of high criticality (frequency greater than or equal to 7), two of average criticality (frequency of 5 or 6), and eight of low criticality (frequency of 3-4). The most related area to usability issues raised in usability rehearsal was the "Info", with twelve issues. The other two areas considered, Interaction and Graphics, were associated with six usability problems each. The four usability problems, classified as high criticality are related to "Info" and "Graph" areas.

4.3 Comparison between Heuristic Evaluation and Interaction Rehearsal Technique

The fourteen (14) usability problems verified in the interaction rehearsal were compared to the twenty-five (25) usability problems observed in the heuristic evaluation, in order to identify if the application had concordance between the two methods. Twelve (12), of the fourteen (14) usability issues of interaction rehearsal, had already been detected in the heuristic evaluation, of which nine (9) were recorded with consent of two or more analysts. Thus, the heuristic evaluation raised thirteen (13) usability issues which were not observed in the test interaction.

It was found that the two (2) usability problems raised just in the interaction rehearsal technique refers to the understanding specific functions that the product has. They are: "Lack of clarity of function 'Delay Start'", which, although doesn't clearly appear as a usability problem in heuristic evaluation, is related to another aspect verified in the same rating on the poor visibility of the operating mode "Delay Start" and "Lack of clarity of 'Spin' function", which was not identified in the inspection without user.

About comparing criticality of usability problems raised from the two methods, it was observed that only one usability problem classified as low criticality in the interaction rehearsal technique obtained in a different heuristic criticality rating (average). The four usability issues of high criticality of interaction rehearsal technique were classified differently in the heuristic evaluation, two low and two average criticality. Attributed to this divergence of criticality to the fact that the heuristic evaluation, criticality was referring to interference in the completeness of the task, the test of interaction, criticality was generated from the frequency of the behavior.

Regarding the time required for the application of this method, for each session it's added a twenty minute period for reviewing post session data. For the tabulation of data, observers demanded eight (8) hours for organization and unification of data, while two analysts have ordered about half a period (four hours). Thus, the interaction rehearsal technique was performed and analyzed in approximately 22 hours in total (not running) demanding: one moderator, one observer, two analysts and ten participants.

It was observed that, both in the heuristic evaluation as in the interaction rehearsal technique, the "Information" area was the most affected, obtaining similar results between the two methods of application. The other two areas (graphic and interaction) also obtained similar results in both assessments. Thus, it is understood that both methods were able to identify the areas of the project development which need improvement. As an illustration of this abstract data comparison, it's given the Figure 1.

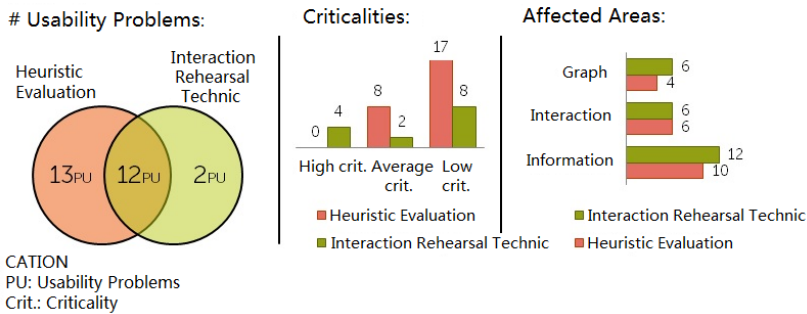


Fig. 1. Comparison of results between usability problems from the heuristic evaluation and the interaction rehearsal technique (Source: the authors)

The interaction rehearsal technique provided data from several users comments, regarding the use, the preference of functions and also the visibility and legibility of elements of the control panel. However, since these comments were not reflected in behavior, they were not counted as usability problems. Thus, the interaction rehearsal technique yielded no data directly related to legibility of elements.

5 Discussion and Final Considerations

From this research, a similarity of results from the application of heuristic rating and interaction rehearsal technique was observed, however, it was also found that the

heuristic evaluation enabled a comprehensive look at the usability problems while testing possible interaction investigate in more detail the positioning of users regarding the identified problems. Thus, it summarizes this scenario with the prospect that the heuristic evaluation allowed visiting different usability problems (horizontal approach), while the interaction rehearsal technique approached fewer problems, but with more details (vertical approach).

Here, representative methods were selected for the use in industry and also in the academy, going against the placement of Roto et. al. (2009) [12] and Dumas & Salzman (2006) [4]. The selection of both methods applied took into consideration the variables presented in ISO 16982:2002 [7], since the product used is an adaptation of the existing system, is a well-understood prototype - given the high fidelity representation of the final product panel - and was able to simulate a product under development. Regarding the presentation of both numeric data such as qualitative commented by Tullis and Albert (2008) [16], it was found that both methods made possible the production of numeric data.

Another aspect observed in the application of both methods is the suggestion of improvements. The heuristic evaluation tends not only to detect problems, but also to suggest ways to fix or improve the interaction as a whole, going against the positioning of Dumas et. al. (2004) [3] as the "positive" reviews of usability. Regarding the heuristics of Nielsen (1994), the golden rules of Shneiderman & Plaisant (2004) [13], the principles of ISO 9241-11 (1008) [8] and the recommendations of Bastien & Scapin (1993) [2], it was found that the analysts used frequently the heuristics and recommendations of Nielsen and Bastien & Scapin.

The heuristic evaluation showed well evaluated for number of generated usability problems, for the analysis of criticality and for the execution time and analysis. The interaction rehearsal technique proved to be more appropriate in capturing actual behavior, investigating user opinions and reviewing and verifying user understanding about the new features. Both methods were not biased by the level of fidelity prototype, indicating that this type of prototype is suitable to conduct evaluations of the control panel during product development.

For the application of the heuristic evaluation, it must be considered the importance of not applying only with a specialist. The worksheet for capturing usability problems proved to be a useful low-cost alternative. The interaction rehearsal technique requires a bigger number of individuals who, even having a value less technical time, shall be accompanied by a moderator and an observer for each session. As this study did not account for costs, because it is voluntary, we consider the "time of application of the method". In this respect, we have that the interaction rehearsal technique demands more time not only for the sessions, but also for the data analysis.

Given the above, we raise the following relationships between these two methods. The heuristic evaluation can be applied at an early stage of evaluation of interface, even before planning a Usability Testing. The results of this Heuristic evaluation would help guide the scenarios, the tasks and the focus of evaluation on a Interaction Rehearsal Technique or Usability Testing. In this view, these two methods are complementary, checking some different problems and reasons levels of criticality from the interaction rehearsal technique. However, if the purpose of the usability study is to develop a broad sweep in finding usability problems, regardless of the

degree of development of the artifact, it is suggested heuristic evaluation as an appropriate alternative, and less costly. However, the application of both methods, especially in the heuristic evaluation, it is important to pay attention to the choice of professionals with experience (theoretical and practical) in the area to perform the activity, since this is a method that directly depends on the expert.

References

1. Annett, J., Stanton, N.A. (eds.): *Task Analysis*. Taylor & Francis, London (2000)
2. Bastien, J.M.C., Scapin, D.L.: *Ergonomic criteria for the evaluation of human-computer interfaces*. Tech. Rep. n. 156. Rocquencourt, France: Institut National de Recherche en Informatique et en Automatique (1993)
3. Dumas, J.S., Molich, R., Jeffries, R.: Describing usability problems: Are we sending the right message? *Interactions* 11(4), 24–29 (2004)
4. Dumas, J.S., Salzman, M.C.: *Usability Assessment Methods*. *Reviews of Human Factors and Ergonomics* 2, 109–140 (2006)
5. Hackos, J.T., Redish, J.C.: *User and task analysis for interface design*. John Wiley & Sons, Inc., Nova York (1998)
6. Hartson, H.R., Andre, T.S., Williges, R.C.: Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction* 10, 145–181 (2003)
7. International Organization for Standardization, ISO TR 16982:2002, *Ergonomics of human-system interaction - Usability methods supporting human-centred design* (2002)
8. International Organization for Standardization, ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability* (1998)
9. Jones, M., Marsden, G.: *Mobile Interaction Design*. John Wiley & Sons Ltd., West Sussex (2006)
10. Karwowski, W., Soares, M.M., Stanton, N.A. (eds.): *Human Factors and Ergonomics in Consumer Product Design, Uses and Applications*. CRC Press, Boca Raton (2011)
11. Maguire, M.: Methods to support human-centred design. *Int. J. Human-Computer Studies* 55, 587–634 (2001)
12. Roto, V., Obrist, M., Mattila, K.V.V.: User experience evaluation methods in academic and industrial contexts. In: *Proceedings of the Workshop on User Experience Evaluation Methods UXEM 2009 (INTERACT 2009)*, ACM, Uppsala (2009)
13. Shneiderman, B., Plaisant, C.: *Designing the UserInterface: Strategies for Effective Human-Computer Interaction*. Addison Wesley, Boston (2004)
14. Stanton, N.A., Young, M.S.: *A Guide to Methodology in Ergonomics: Designing for human use*. Taylor & Francis, Londres (1999)
15. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: *Human factors methods: a practical guide for engineering and design*. Ashgate, Cornwall (2005)
16. Tullis, T., Albert, B.: *Measuring the user experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann, San Francisco (2008)