ITEA Office
High Tech Campus 69 - 3
5656 AG Eindhoven
The Netherlands
T + 31 88 003 6136
E info@itea3.org
W www.itea3.org

ITEA 3 is a EUREKA strategic ICT cluster programme

D1.1.1 Evaluation Methods & Tools

ModelWriter

Text & Model-Synchronized Document Engineering Platform

Work Package: WP1
Task: 1.1 Industrial Use Cases

Edited by: Ersan Gürdoğan <ersan@hisbim.com> (HISBIM)

Date: 08-Sep-2015
Version: 1.0.0

Document History

Version	Author(s)	Date	Remarks
1.0.0	Ersan Gürdoğan	08-Sep-2015	Initial Release

Table of Contents

1. Introduction

Evaluation should show:
1) What actually occurred?
2) Whether it had an impact, expected or unexpected, and
3) What links exist between a program and its observed impacts?

First, knowing what actually occurred in the course of the program means knowing who did what to whom. Information on what happened is captured in what are known as process measures, outputs or, sometimes, widget counting. These measures are not sufficient to determine whether or not the program had an impact, but without them it isn't clear whether the program was carried out according to plan. Evaluators need evidence that the program was delivered in order to make responsible statements about causation. Second, the evaluator needs to know the impacts of the program, both those the program is expected to achieve and also unexpected positive or negative impacts. The expected impacts are codified in statements of program objectives. An evaluator often takes baseline measurements to document the situation before an intervention, and then takes the same measurements after an intervention to assess impact. However, this technique only works for expected impacts. The unintended impacts are more challenging to document, because there are usually no baseline measures for them. Yet, identifying unintended impacts can be important in understanding all of the impacts of a program. Third, the evaluator needs to make responsible judgments about the links between a program intervention and its observed impacts. While a certain amount of conjecture enters into making these links, increased knowledge about the program as it was delivered, and its impacts, leads to stronger conjecture. Most traditional evaluation designs use quantitative measures, collected over a sample of the population, to document these three stages. However, there are times when this sort of evaluation design does not work as effectively as a case study evaluation. This guide is designed to help evaluators assess whether or not a case study is a useful evaluation tool for a given project, and if so, this guide explains how to do a good case study evaluation. Like any other evaluation design, the case study should suit the project to which it is applied and must be well executed for maximum benefit. Methods & Tools should answer the questions below?

- What are we trying to achieve?
- Is the role of the group to advise or steer the project?
- Who will draft the plan (or protocol)?
- Who and how will the data be collected?
- Who and how will the data be analysed?
- Who will write up the final report?
- What is the time frame?

Usability evaluation with the Use Case Evaluation (UCE) method consists of three overall activities, see Fig. 1:

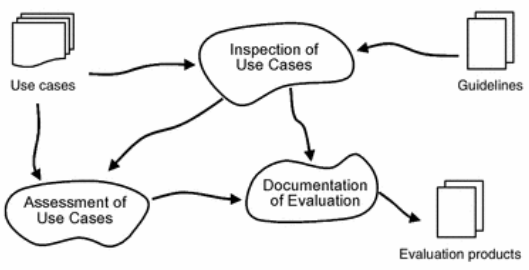

Fig. 1. Overview of Activities and Materials in Use Case Evaluation (UCE)

2. Theoretical Approaches

2.1. A brief explanation of some methodologies

2.1.1. Evaluation

An in-depth study which takes place at a discrete point in time, and in which recognised research procedures are used in a systematic and analytically defensible fashion to form a judgement on the value of an intervention (UKES 2011).

2.1.2. Before and After Design

An example of a quasi-experimental design which simply compares the situation after the intervention with the situation beforehand and attributes any difference to the effects of the intervention. A particular weakness of this design is the possibility that something else besides the intervention accounts for all or part of the observed difference over time.

2.1.3. Case Study

A data collection technique involving the examination of a limited number of specific cases or projects which the evaluator anticipates will be revealing about the programme as a whole. Case studies tend to be appropriate where it is extremely difficult to choose a sample large enough to be statistically generalizable to the population as a whole; where generalisation is not important; where in-depth, usually descriptive data is required; and where the cases or projects to be studied are likely to be quite complex.

2.1.4. Comparative Change

An example of a quasi-experimental design in which any known or recognisable difference between the programme and control groups is taken into account in the statistical analysis. The problems with this design are, firstly, that there may be some other factor which explains some or all of the variation in the intervention and in the observed effects, and, secondly, that there may be initial differences between the programme and control groups which have an influence on observed effects and which can therefore become confounded with the influence of the programme on these effects (selection bias).

2.1.5. Cost Benefit Analysis

A judgemental technique in which a researcher compares all social and other costs and benefits of a programme with a view to determining whether the benefits exceed the costs, and if so by how much. Social costs and social benefits usually have to be measured by some indirect means and converted into monetary values so that a comparison can be made with other costs and benefits. Furthermore, it may not be appropriate to use prevailing market prices. Consider a situation of very high unemployment. In this case, the real cost of labour may be much lower than the prevailing market wage. The opportunity cost (the next best use of the otherwise unemployed workers had the project not gone ahead) is lower than the prevailing wage rate, and this low opportunity cost has to be represented by a shadow price which has to be derived somehow.

2.1.6. Cost Effectiveness Analysis

A judgmental technique in which the researcher quantifies the costs and benefits associated with a programme on the basis of the same principles which apply to cost-benefit analysis, but there is no requirement to transfer benefits into common monetary units.

2.1.7. Criterion Population Design

An example of a quasi-experimental design, which attempts to improve on the comparative change design. In the latter, the programme and control groups are two distinct groups drawn from a hypothetical larger population. In the criterion population design, however, the hypothetical population is identified and used for the comparison group. In this case, the possibility of selection bias is confined to just one group - the programme group. This design is particularly appropriate where the evaluator cannot easily create a control group but does have access to information about the larger population from which the programme group is drawn.

2.1.8. Delphi Technique

A method for structuring a group communication process, allowing a group of individuals, as a whole, to deal with a complex problem. Experts are consulted separately in a number of different rounds. In each successive round, each individual is told the views of the other experts in the previous round. This technique can be used to arrive at a consensus, or at least to reduce disagreement.

2.1.9. Ex Post Facto Design

This is an example of a descriptive design. It can be used where the evaluator cannot select who is to be exposed to the programme, and to what degree. These designs have been used to examine interventions with universal coverage. Also known as causal comparative research.

2.1.10. Ideal Experimental Design

A theoretical way of determining the impact of an intervention. It involves comparing two groups which are identical in all respects except one: exposure to the intervention. Differences between the group which has been exposed (the programme group) and the group which has not (the control group) are then attributable to the intervention. In the real world, this design does not exist since we can never be absolutely certain that the two groups are identical in all other respects. The potential non-equivalence of the two groups weakens the validity of any causal inference about the intervention.

2.1.11. Interrupted Time Series

An example of a quasi-experimental design. This involves obtaining several measurements over time both before and after exposure to a programme in order to create a timed series of observations. It is purported to be an improvement on the before and after design.

2.1.12. Logic Models

Logic modeling is a tool that can be useful in the development of monitoring and evaluation plans as they help to identify short-, medium- and long-term outcomes that are linked to the key activities of a programme or strategy. Typically, logic models are developed in collaboration with key stakeholders involved in strategy development and delivery.

2.1.13. Literature Review

A data collection technique which enables the evaluator to make the best use of previous work in the field under investigation and thus to learn from the experiences and findings of those who have carried out similar or related work in the past. There are two types of documents that can be used in a literature search. Firstly, there are published papers, reports and books prepared by academics, experts and official organizations. Secondly, there are specific studies in the area, including past evaluations and policy documents. These can be harder to source and are often referred to as 'grey literature'.

2.1.14. Natural Observations

A data collection technique in which the evaluator makes on-site visits to locations where the intervention is in operation and directly observes what is happening. Observational data can be used to describe the setting of the intervention, the activities which take place in the setting, the individuals who participate in these activities (who may or may not be aware that they are being observed), and the meaning of these activities to the individuals. This form of data collection is particularly vulnerable to the Hawthorne effect.

2.1.15. Participatory Action Research

Participatory Action Research involves individuals and groups researching their own personal beings, social settings and experiences. They reflect on their values, shared realities, collective meanings, needs and goals. Knowledge is generated and power is regained through deliberate actions that nurture, empower and liberate individuals and groups. The researcher works in partnership with the participants throughout the research purpose. PAR can be time consuming because sometimes delay can occur when researchers from outside the community and community members need to negotiate phases in the research. Research using PAR should ideally describe the negotiation processes used.

2.1.16. Performance Audit

Conceptually closer to evaluation than traditional audit, performance audit is strongly concerned with questions of efficiency (of an intervention's direct outputs) and good management. Performance audit and evaluation share the same aim of improving the quality of programmes, but evaluation goes much further. It also looks at issues such as sustainability, relevance and the longer-term consequences of a programme.

2.1.17. Plan, Do, Study, Act Cycle (PDSA)

The plan, do, study, act (PDSA) cycles can be used to test an idea by temporarily trialling a change and assessing its impact. This approach is unusual in a healthcare setting because traditionally, new ideas are often introduced without sufficient testing. The four stages of the PDSA cycle are Plan - the change to be tested or implemented; Do - carry out the test or change; Study - data before and after the change and reflect on what was learned; Act - plan the next change cycle or full implementation.

2.1.18. Quasi Experimental Design

A quasi-experiment is an observational study in which the subjects to be observed are not randomly assigned to different groups in order to measure outcomes, as in a randomized experiment, but grouped according to a characteristic that they already possess.

2.1.19. Realist Evaluation

The term ‘realist evaluation’ is drawn from Pawson and Tilley’s seminal work (1997). The key to this methodology is that Context + Mechanism = Outcome. Therefore the task of evaluators is to learn more about ‘what works for whom’, ‘in which contexts particular programs do and don’t work’, and ‘what mechanisms are triggered by what programs in what contexts [2].

2.1.20. Regression Analysis

A statistical technique which can be used to establish the significance of any correlation (association) between variables of interest, e.g. the gender of a long-term unemployed worker and the amount of time before he or she finds a new job after a training programme. In regression analysis, there is an attempt to establish whether the variation in one variable (known as the dependent variable) can be explained in terms of the variation in one or more independent variables. The dependent variable is often quantitative. Special techniques are available, however, to deal with situations in which the dependent variable is qualitative.

2.1.21. Research Synthesis

An overview of the current state of knowledge about a socio-economic problem and about remedies through public policy, which is undertaken before an evaluation. This knowledge can be obtained from professional literature, media articles, and administrative data, monitoring reports or published statistics. Preparing a research synthesis is often helpful prior to launching an evaluation. By listing the information that is available and comparing it with the needs ensuing from the analytical agenda, the research synthesis will point to the principal information gaps which, in turn, set the data collection and analysis tasks to be undertaken by the evaluation. Reviews of literature can also be a data collection technique in the conduct of an evaluation.

2.1.22. Thematic Evaluation

An evaluation which focuses on one or more themes which are common to several different interventions (programmes or other activities), for example, effects on the environment or on small and medium-sized organisations.

2.1.23. True Experimental Design

The best real world approximations to the ideal experimental design, in which the evaluator tries to ensure the initial equivalence of the programme and control groups by creating them beforehand through random assignment. Although causal inference based on such designs is usually very strong, true experimental designs are difficult to administer and implement. Also referred to as "randomised experimental design.’’

3. Data Collection and Analysis

3.1. Collecting the Data

Once you have decided what you want to measure or find out, you must decide how you are going to do it. Two possibilities are available – using existing data (Secondary data) or collecting new data (Primary data). Existing data can give information in very large numbers, usually cheaply and quickly and with minimal disruption to patient care. However it is important to ensure that you collect the right data to address the objectives, the data is in the right form to carry out an appropriate analysis and you collect enough data to answer the evaluation questions. This might mean it is necessary to create new data. To collect new data you will need to consider which data collection tools (such as surveys, interviews, focus groups) you require to ensure you acquire the data you need to achieve your goals.

3.1.1. Data collection tools

It is imperative that the data collection tool you choose actually measures what you intend it to. The term validity is used to describe how well the tool measures the 'truth'. For example, to what extent does a self-reported food frequency questionnaire reflect the actual dietary intake of the individual completing the questionnaire? Probably not at all – so as an evaluator you may feel that, in these circumstances, this type of data collection tool is not as valid as another tool might be. There are a large number of validated tools already available, these include patient reported outcome measures (PROMs). It is worth exploring what is available that may be relevant to use. Another factor to consider is reliability e.g. how likely it is that a measurement instrument will measure the same thing each time it is used? If the measurement cannot be repeated with any consistency you may decide that it is unreliable. For these reasons it is best to test or pilot any data collection tools that you design, before you use them with your chosen sample of participants. Some of the tools for data collection are listed below.

3.1.2. Examples of data collection tools

- Focus groups: A group of 6 -10 people, who participate in an organised, guided discussion, led by a facilitator, with the purpose of gaining their views and experiences of a particular topic or activity. Qualitative data is produced by drawing on the participant's attitudes, feelings, beliefs, experiences and reactions demonstrated through the group interaction. Essentially, the aim is to get at how people feel and what their thoughts and perceptions are. Please ask us for more advice and information on how to plan and undertake a focus group.
- Interviews: An interview is a discussion between an interviewer and a participant. This can occur face to face, or by telephone, usually using list of questions and prompts. Interviews may be structured (with fixed, but usually open ended questions) or semi-structured (where discussion can be more flexible).
- Questionnaires: These can be a cheap and quick way of seeking the views of cases and system but without careful design, can lead to misinterpretation, misrepresentation of views and a low response rates. There are several types of questions, which all have strengths and weaknesses depending upon the context in which they are used. Designing a good questionnaire is a skilled job, and it is best to seek advice. We can help with advice and information on how to design and use a questionnaire.
- Document analysis: The analysis of a written document may prove to be a useful addition to other data collection methods. The types of document you might wish to review include

patient records, local policy or protocols, minutes of meetings and practitioner reports or statements. What is important is that you review these documents in a consistent way. Content analysis, dividing the data into categories or themes, is an appropriate way to do this but can take considerable time and skill to achieve successfully.

- Monitoring forms: Sometimes it is necessary to devise a monitoring form to collect new data. This can be time consuming for busy practitioners, and should be avoided if possible, unless it can be incorporated into an existing system of record keeping. This is much easier to achieve if the evaluation process is started at the outset of any new service development and is part of the service design and planning. There are information governance rules about accessing patient records.
- Literature Review: A data collection technique which enables the evaluator to make the best use of previous work in the field under investigation and thus to learn from the experiences and findings of those who have carried out similar or related work in the past. There are two types of documents that can be used in a literature search. Firstly, there are published papers, reports and books prepared by academics, experts and official organizations. Secondly, there are specific studies in the area, including past evaluations and policy documents. These can be harder to source and are often referred to as 'grey literature'.

3.2. Analysing Your Data

Data analysis is a process by which the data you have collected is transformed into meaningful and useful information. It can be a very complex process, with many possible approaches that are specific to the data collection methods used. The most common analysis of qualitative data is observer impression. That is, the expert or layman observers examine the data, interpret it via forming an impression and report their impression in a structured form. The analysis of quantitative data is also structured using a validated statistical tool. Anything other than simple counting of the number of occurrences will require the input of an experienced researcher or statistician.

4. Use Case Evaluation

Usability evaluation with the Use Case Evaluation (UCE) method consists of three overall activities: (1) Inspection of use cases, (2) Assessment of use cases, and (3) Documentation of evaluation. The input for inspection is a collection of use cases describing the use of the system under development and a brief description of the use context for the system. We recommended the fully dressed form of use case description proposed by Cockburn [5]. With fully dressed use cases, evaluators receive as much information as possible with a use case. In addition, a list of guidelines is required to assist the inspection. The evaluation product is an assessment of the usability of the system expressed as a list of usability problems. It may also include an assessment of the quality of the use cases. The evaluation product is subsequently fed back into the interaction design activity [1].

4.1. Guidelines for Inspection

UCE provides a set of usability guidelines that we suggest as being particularly apt for use case inspection, see Table 1. The guidelines were originally derived from previous research. Our aim has not been to make the guidelines non-overlapping, but merely to provide rich and varied support for identifying usability problems. UCE is based mainly on heuristics introduced as part of Heuristic Evaluation [3,4]. These heuristics were chosen because they have shown to be applicable across a wide range of contexts, and because they are among the most extensively validated inspection guidelines. Some of these heuristics do, however, concern details of the user interface design that will typically not be specified at the time use case evaluation is likely, for instance the heuristics ‘Aesthetic and minimalist design’ and ‘Help and documentation’. Therefore, they have been excluded from our list. We have supplemented the heuristics with guidelines from other methods that help focus on other usability concerns, or provide better examples or more detail (see Table 1). Guideline 9 is based on a principle from Cognitive Dimensions [5], somewhat similar to an Ergonomic Criteria [5] on designing for low workload. We also used Cognitive Dimensions [5] to add guideline 10 on premature commitment. Premature commitment occurs when software requires users to do an action or supply information that they are not ready to do or supply. It is related to a principle in Cognitive Walkthrough [6]. Guideline 11 was motivated by the idea that early evaluation on use cases should help establish the utility of the proposed functionality. In particular, we wanted to allow evaluators to focus on issues like task relevance and missing functionality. It has been suggested that evaluators might be less attentive to utility issues than to issues of convenience and to surface-level interface issues [7].

4.2. Inspection of Use

The inspection activity is conducted by one or more evaluators. It involves the following two steps: Brainstorm and Systematic Inspection. In the first step, Brainstorm, the evaluator goes through the use cases one by one without following any systematic procedure. The evaluator notes the usability problems that may be predicted from the use cases. In the second step, Systematic Inspection, the evaluator employs a structured procedure for inspection of use cases. This procedure follows the one proposed in the early papers on Heuristic Evaluation [3,4]. It is supported with the guidelines presented above. The inspection typically lasts one to two hours. The use cases are inspected one by one. For each use case, the evaluator tries to predict the usability problems a user will experience while carrying out the use case. In doing this, the evaluator employs the guidelines. The aim is to couple the ideals of the guidelines to the use

cases, trying to see similarities and cases where a guideline may be breached. As recommended for heuristic evaluation, it is fruitful to go over all use cases at least twice. The evaluator produces a list of usability problems; if more than one evaluator inspects the interface the evaluators' lists are merged into one joint list.

Guidelines for Use Case Evaluation No Guideline Explanation Sources

1. Visibility of system status
   The system should always keep users informed about what is going on, through appropriate feedback within reasonable time [4]
2. Match between system and the real world
   The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system oriented terms. Follow real-world conventions and make information appear in a natural order. [4]
3. User control and freedom
   Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo. [4]
4. Consistency and standards
   Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions. [4]
5. Error prevention
   Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Eliminate error-prone conditions or handle them gracefully. [4]
6. Recognition rather than recall minimize the user's memory load by making objects, actions, and options visible.
   The user should not have to remember information from one part of the dialogue to another. [4]
7. Flexibility and efficiency of use
   Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. [4]
8. Help users recognize, diagnose, and recover from errors
   Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution. [4]
9. Avoid hard mental operations and lower workload
10. Do not force the user into hard mental operations and keep the user's workload at a minimum. [4,5]
11. Avoid forcing the user to premature commitment
12. Do not force the user to perform a particular task or decision until it is needed. Will the user know why something must be done? [4,6]
13. Provide functions that are of utility to the user
14. Consider whether the functionality described is likely to be useful to users and whether functions/data are missing.

4.3. Assessment of Use Cases

A secondary activity is to assess the quality of the use cases. In contrast to heuristic evaluation of a fully functional system, there is likely to be several cases where a use case does not give an evaluator sufficient information to decide whether or not a guideline is breached. In those cases, supplementing information may need to be provided or the evaluator may simply express why

something cannot be properly analyzed. Sometimes, the use cases will not be specific enough to allow evaluation. In other cases, they will be too specific, e.g., by specifying user interaction details that can be decided later in the development process. Thus the outcome is an assessment of each use case, which emphasizes how useful the use case is for inspection. In practice, this activity is done in parallel with the primary evaluation activity.

4.4. Documentation of Evaluation

In this activity, the results are compiled into a coherent evaluation product that may be fed back into interaction design. The main content of the documentation is the list of usability problems. This list describes problems that the evaluators expect a prospective user will face when using the system. Each of these predicted usability problems should include a clear reason why it is perceived to be a problem. For example, the reason could be a reference to a guideline. As with heuristic evaluation, evaluators should describe only one problem at a time and be as specific as possible. Finally, we suggest that evaluators note ideas for improving the system or designing it differently. A supplementary evaluation product is the assessment of each use case.

5. Empirical Study

The method described above needs to be assessed with respect to its usefulness for use case evaluation. Thus we conducted an empirical study aimed at (a) comparing usability problems identified with the UCE method to a set of problems discovered with think-aloud testing and (b) understanding what problems evaluators experience in using UCE and which guidelines that are particularly useful for inspecting use cases. The rationale for (a) is to investigate if the UCE method find problems that appear in a final version of a system. The rationale for (b) is to obtain input that may be used to improve early us case inspection with the UCE method.

5.1. Evaluators

We had four evaluators use the UCE method to evaluate use cases; two of them are among the authors of this paper. We chose to use four evaluators because it conforms to the often recommended, though controversial, number of evaluators for inspection, and because it seemed a realistic number of evaluators to find when bringing UCE to practical use. The evaluators were all experienced with usability work (from two to eight years of experience) and all had conducted several think aloud tests and usability inspections. Two of the evaluators had a master degree in computer science and were working in industry, one was a PhD student working with usability evaluation, and one was associate professor in human-computer interaction.

5.2. System and Use Cases for Evaluation

UC-TR-01	Documents of Quality Assurance Department
Version	V1.0.0 dated 15-Nov-2014
Description	To create faster and more accurate the forms that are used in quality control progress and trigger related forms (re-work form triggering, revision needed, approved, rejected etc.)
Actors	Quality Managers, Quality Measurement Specialists, Quality Control Personnel, Quality Auditors, Production Crews
Assumptions	Quality control measurement units are standard Rejected and Approved products forms are standard Quality Certification standards are always applied
Steps	Products information forms are created according to product information comes from ModelWriter system Products are measured by quality control department authorized personnel according to standards To decide product is standard or not according to measurement report Product would be sent rejected products section if its measures are out of standards ModelWriter system is created Rejected Product Quality Form Else Approved Product Quality Form is created by ModelWriter System Approved products are sent to warehouse as accepted products
Alternatives	If rejected products' measurement out of range in rework standards, it means cannot be applied rework on this product, it returns to scrap. Then it should be sent wasteland.

Issues	Standardizing form designs is hard because of all customers have their own unique reporting tools/formats

UC-TR-02	Non-Disclosure Agreements
Version	V1.0.0 dated 15-Nov-2014
Description	Non-Disclosure Agreement (NDA), also known as a confidentiality agreement (CA), confidential disclosure agreement (CDA), proprietary information agreement (PIA), or secrecy agreement (SA), is a legal contract between at least two parties that outlines confidential material, knowledge, or information that the parties wish to share with one another for certain purposes, but wish to restrict access to or by third parties. It is a contract through which the parties agree not to disclose information covered by the agreement. An NDA creates a confidential relationship between the parties to protect any type of confidential and proprietary information or trade secrets. As such, an NDA protects nonpublic business information. NDAs are commonly signed when two companies, individuals, or other entities (such as partnerships, societies, etc.) are considering doing business and need to understand the processes used in each other's business for the purpose of evaluating the potential business relationship. NDAs can be "mutual", meaning both parties are restricted in their use of the materials provided, or they can restrict the use of material by a single party. It is also possible for an employee to sign an NDA or NDA-like agreement with an employer. In fact, some employment agreements will include a clause restricting employees' use and dissemination of company-owned confidential information.
Actors	Responsible/Authorized personnel in both parties.
Assumptions	Agreements are prepared and written according to European Business Law.
Steps	To define both partied who would sign the NDA To write items of agreement according to scope of NDA Reviewing agreement by decision maker then getting approval Sharing NDA each other Send feedback to ModelWriter system if any change apply on NDA Sahepen final version of NDA then signing by both parties
Alternatives	After sharing NDA, both parties sign without any change and no feedback. Cancellation of NDA
Issues	Different laws could be applied in out of European Union

6. References

1. Hornbæk, K., Stage, J.: The Interplay between Usability Evaluation and User Interaction Design. International Journal of Human Computer Interaction 21(2), 117–123 (2006)
2. http://www.evaluationontario.ca/pdf/2004_NickTilley_CESO.ppt
3. Molich, R., Nielsen, J.: improving a human-computer dialogue. Communications of the ACM 33(3), 338–348 (1990)
4. Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R. (eds.) usability inspection methods, pp. 25–62. John Wiley & Sons, Chichester (1994)
5. Green, T.R.G.: Cognitive Dimensions of Notations. In: Proceedings of People and Computers V, pp. 443-460 (1989)
6. Wharton, C., Rieman, J., Lewis, C., Polson, P.: The cognitive walkthrough method: a practitioner's guide. In: Nielsen, J., Mack, R.L. (eds.) Usability inspection methods, pp. 105–140. John Wiley & Sons, Chichester (1994)
7. Nørgaard, M., Hornbæk, K.: What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. In: Proceedings of DIS2006 (2006)
8. http://www.case.edu/affil/healthpromotion/ProgramEvaluation.pdf
9. http://link.springer.com/chapter/10.1007%2F978-3-540-74796-3_58#page-1