

---

# Chapter 14. Usability Analysis

## Table of Contents

The Role of Usability Analysis .....	1
Usability analysis in the development lifecycle .....	1
Analytical techniques .....	1
Predictive Modelling .....	2
The GOMS model .....	2
The Keystroke level model .....	2
Cognitive walkthroughs .....	3
Empirical Analysis .....	3
User Testing .....	3
Field Studies .....	3
Heuristic Evaluation .....	4

## The Role of Usability Analysis

Analysis aims at assessing how well an idea, design, solution or approach does its job. For most IT research, this cannot easily be done with a few simple calculations, but requires designing, developing and testing an experimental system, and then interpreting the results. The aim of an interface evaluation (or usability analysis) is to understand user needs better, to get feedback for improving a design, to compare alternatives in order to choose between them, or to determine how well a design meets users needs and how much they like it.

Usability analysis can be qualitative or quantitative, and most often encompasses both kinds of assessment. The main qualitative products of usability evaluation are descriptive reports, quotes, lists of problems (possibly with suggested solutions) and anecdotes; quantitative results are performance measures and error rates associated with tasks (or calculated from theoretical models, in the case of predictive evaluation).

## Usability analysis in the development lifecycle

Evaluations done during design are known as formative evaluations; those done on the finished product are called summative evaluations. Typically a small number of users comment on early prototypes; this is fed into subsequent redesign and followed by more formal evaluation. Effective research and development requires knowing how to evaluate systems at different stages of the design-evaluate-redesign cycle. With each evaluation, both designers and users obtain a better grasp of what is required. Evaluation is a good way of ensuring that users are involved in system design; it is also cheaper to fix problems then, rather than after deployment.

Quick-and-dirty evaluations are relatively informal and give rapid feedback of a qualitative nature from users or usability consultants. They are useful initially as a short pilot study, followed by one or more iterations of a more thorough design-evaluate-redesign cycle involving modelling, walkthroughs, expert evaluation, user testing, interviews, questionnaires and/or observations. Quick-and-dirty evaluations are also useful in later iterations to check out specific details like the suitability of a particular icon.

## Analytical techniques

Two main types of assessment are analytical (through modeling or simulation) and empirical (by building and testing a prototype). Analytical evaluation methods in user studies are:

1. predictive evaluation using a specific model of interaction
2. cognitive walkthroughs, analysing a task step by step

Empirical methods are:

1. user testing in the laboratory
2. field studies
3. heuristic evaluation by experts

## Predictive Modelling

In predictive evaluation, a model of the actions required to perform some task is developed first, and then these individual actions are analysed to evaluate the task as a whole. A task to be analysed is called a benchmark, and forms the basis of comparisons between systems, or between alternative versions of a system. A few models of human-computer interaction exist which can be used to predict effectiveness. The most well-known is the GOMS model, and a particular variant of it called the keystroke-level model.

## The GOMS model

The GOMS model [Card 83] of predictive user modelling stands for Goals Operators Methods Selection rules. It is a model of the knowledge, thinking and actions employed by users of a system.

- Goals are states a user wants to achieve. They are often decomposed into sub-goals.
- Operators are the thinking and actions needed to achieve those goals.
- Methods are a series of steps to accomplish a goal, i.e. a procedure or recipe.
- Selection rules are used to choose between different ways of achieving the same goal.

With this approach, a set task is modelled in terms of sub-goals, selection rules, methods and their operators. This breakdown is then used to analyse that task. The GOMS model is most effective for predicting performance times. It is also useful for comparing alternative methods of performing the same task, since different methods can be enumerated along with their predicted performance times.

There are several reports where using the GOMS model was very successful, as the detail breakdown of tasks showed where problems would arise and also precisely what was causing them (e.g. in [Gray 93]). Shortcomings of the GOMS model however include its inability to model anything but expert users in laboratory situations working on a limited set of tasks. Tasks that are complex, take a long time or have a wide variety of options are not suited to GOMS modeling. It also does not take into account the real-world environment in which the system will be used, where distractions, pauses and changes frequently occur. GOMS models are thus very helpful in predicting performance of experienced users do routine tasks, but are limited in scope where other kinds of systems and users are involved.

### References:

- S.K.Card, T.P. Moran & A. Newell. The Psychology of Human-Computer Interaction. Lawrence Erlbaum Assoc: 1983.
- W.D. Gray, B.E. John & M.E. Atwood. Project Ernestine: validating a GOMS analysis for predicting and explaining real-world performance. Human-Computer Interaction 8(3), pp. 237-309: 1993.

## The Keystroke level model

This model is based on the average amount of time taken to perform common user tasks such as pressing a key or deciding what to do. The predicted time of accomplishing some task is computed by listing the actions it requires, and adding up the times for each of those steps. This enables systems to

be compared with each other, new designs to be compared against existing ones, alternative methods within a single system to be compared, etc.

A problem with this model is that thinking and decision-making times can vary greatly depending on the individual and the situation. Another is that it is not always clear where such mental activities need to be factored in (where will users need to think or spend time choosing between options?) and it is important to be consistent if comparisons are to be correct. In keystroke-level analysis, mental preparation steps should generally be inserted before user operations, but this depends on the operations involved; experience with the method is needed to introduce mental operations in an appropriate way. Some rules for doing so are proposed in [Card 83].

**References:** S.K.Card, T.P. Moran & A. Newell. The Psychology of Human-Computer Interaction. Lawrence Erlbaum Assoc: 1983.

## Cognitive walkthroughs

Cognitive walkthroughs are used to model how usable a system will be to inexperienced or new users. That is, they provide a means of evaluating how successfully exploratory learning can be done on that system. They are therefore complementary to GOMS models, which focus on experienced usage.

In a walkthrough method, designers walk through a task using the interface, and analyse the process step by step. Where there are many alternative ways of executing the task, it is sometimes adequate to choose one and sometimes necessary to model a few of these possibilities. In some application areas it is necessary to study users in the field in order to understand the tasks they perform, before walkthroughs can be attempted.

Walkthroughs are concerned mostly with error handling what errors are possible, how discernible are they, how easy is error recovery, what are the chances of novice users successfully completing a task, etc. - rather than with performance issues. Cognitive walkthroughs are time-consuming, as all common tasks need to be stepped through and each analysed for a range of different usage conditions.

## Empirical Analysis

Empirical user studies involve testing a prototype system using real users in a laboratory experiment, or by conducting a field study, or through experts applying heuristic evaluation.

It is essential to have at least one pilot study before conducting a test, so that problems with instructions, questionnaires and data analysis can be detected - before time and money is wasted on a full-scale experiment. Pilot studies also enable researchers to improve their skills in observing, interviewing and conducting field studies, and to enhance the prototype itself - all of which can lead to substantial improvements in the quality of the final evaluation.

## User Testing

In usability testing, researchers construct a set of tasks for users to perform; measure the speed and accuracy of the people who execute these, count the number of problems they encounter, and note problems encountered, navigation paths used and any interesting or unanticipated events. The test is conducted as a scientific experiment under strictly controlled conditions. Sometimes performance is evaluated against optimal and minimally acceptable times, at other times the experiment is exploratory in order to improve understanding of user needs and behaviour. Most commonly, two systems are compared to each other by comparing the performance of a group using the one against that of a group using the other. Interviews and/or questionnaires typically follow the test.

## Field Studies

There is a growing trend towards evaluating software in the environment in which it will be used, rather than in the laboratory. This enables its usage in natural settings to be observed, recorded and analysed.

Field studies are essential for critical applications involving health, security or business success, and are strongly recommended for collaborative systems. They are far lengthier than laboratory experiments because time is needed for the users to be trained and to adapt to the system in their environment, for databases to be loaded, and for evaluators to obtain sufficient data.

In a field study the aspects to investigate are often finalised only after the system is deployed, once the characteristics of usage in that application context become clear. The system should be complete or nearly so, in contrast with laboratory experiments where simplified prototypes are best in the early stages of the project. Analysis of field studies should emphasise comparison of performance before and after the introduction of the system. One of the most valuable results of a field study is often the insight that the research team gains into the application environment and its demands on the system.

## Heuristic Evaluation

In heuristic evaluation one or more user interface experts evaluate a design to identify problems. A collection of design rules or heuristics is used to guide them. In the sense that expert opinions are obtained, they are the usability equivalent of software inspections and code reviews.

Usability heuristics proposed by Nielsen and Molich [Nielsen & Molich 89] include:

- Simple, natural language
- Consistency
- Short cuts provided
- Exits provided
- Feedback provided
- Good error messages
- Minimal memory load

It is best to complement user testing in the field or laboratory with studies that are not dependant on having a representative sample of the user population. Heuristic evaluation can be applied in situations where cognitive walkthroughs and GOMS analyses are impractical (i.e. for studying situations other than the two extreme cases of novice and of routine, experienced usage respectively). Another advantage of heuristic evaluation is that it is performed by outsiders, not the system developers themselves. Unlike other methods, it requires no advance planning. On the other hand, it is less repeatable than other analysis methods, which is why it is preferable to have a number of expert evaluators and to then aggregate the problems identified.

**References:** J. Nielsen & R. Molich. Teaching user interface design based on usability engineering. ACM SIGCHI Bulletin 21(1), pp. 44-48: 1989.