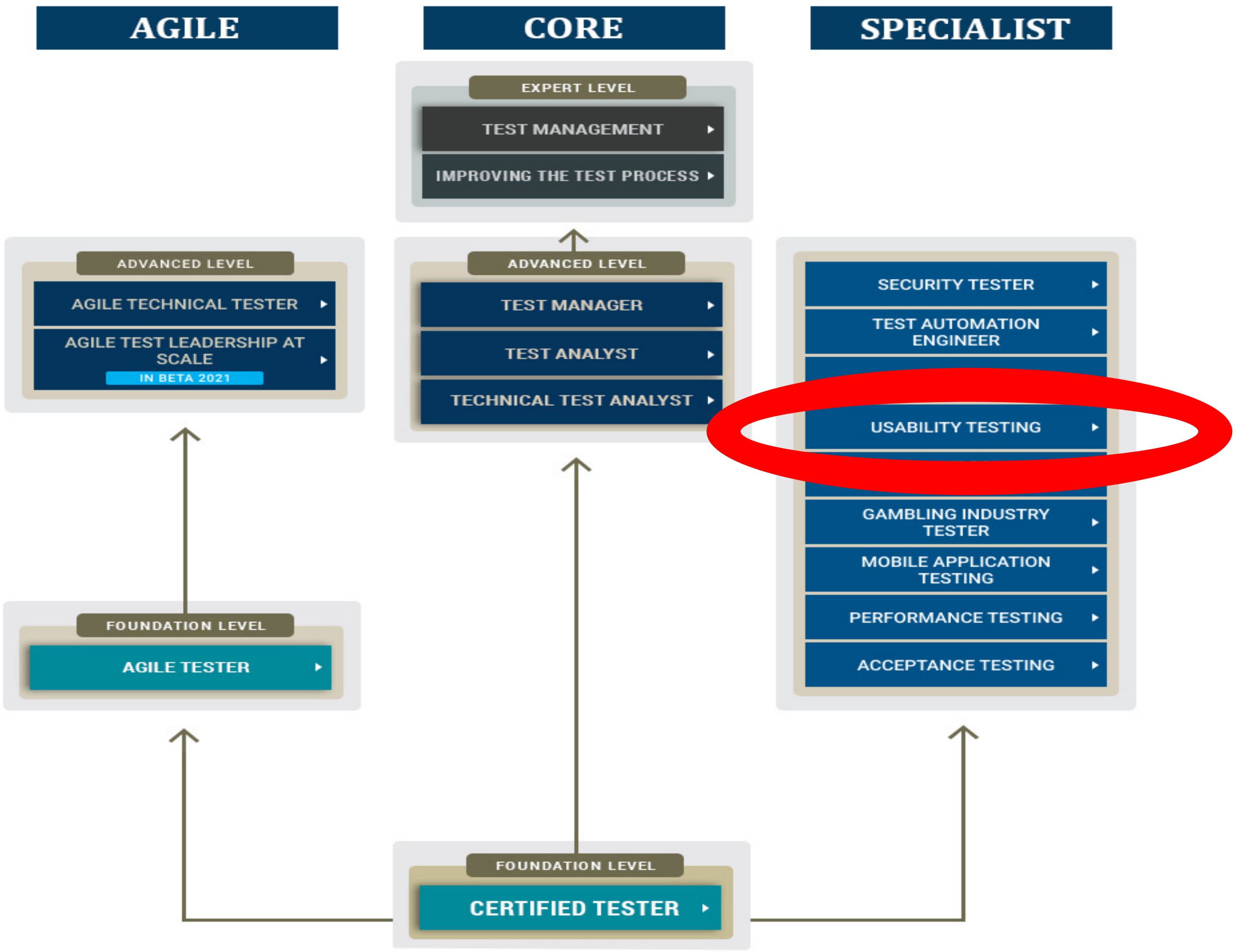
Usability Testing

A Collection of Fundamental Contents Extracted from Dr. Jeff Winter’s Usability Testing Lecture Slides

ISTQB
International Software
Testing Qualifications Board



AGILE

CORE

SPECIALIST

EXPERT LEVEL

TEST MANAGEMENT ▸

IMPROVING THE TEST PROCESS ▸

ADVANCED LEVEL

AGILE TECHNICAL TESTER ▸

AGILE TEST LEADERSHIP AT SCALE ▸
IN BETA 2021

ADVANCED LEVEL

TEST MANAGER ▸

TEST ANALYST ▸

TECHNICAL TEST ANALYST ▸

SECURITY TESTER ▸

TEST AUTOMATION ENGINEER ▸

USABILITY TESTING ▸

GAMBLING INDUSTRY TESTER ▸

MOBILE APPLICATION TESTING ▸

PERFORMANCE TESTING ▸

ACCEPTANCE TESTING ▸

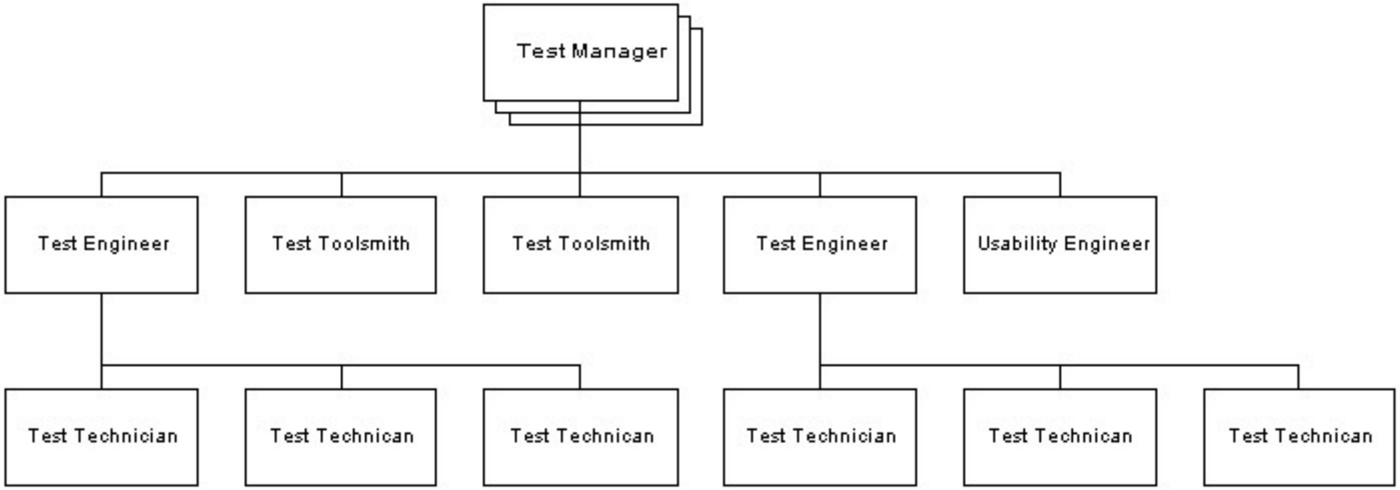
FOUNDATION LEVEL

AGILE TESTER ▸

FOUNDATION LEVEL

CERTIFIED TESTER ▸

Case Study: Positions, Organization, Project



Test Manager

Test Engineer

Test Toolsmith

Test Toolsmith

Test Engineer

Usability Engineer

Test Technician

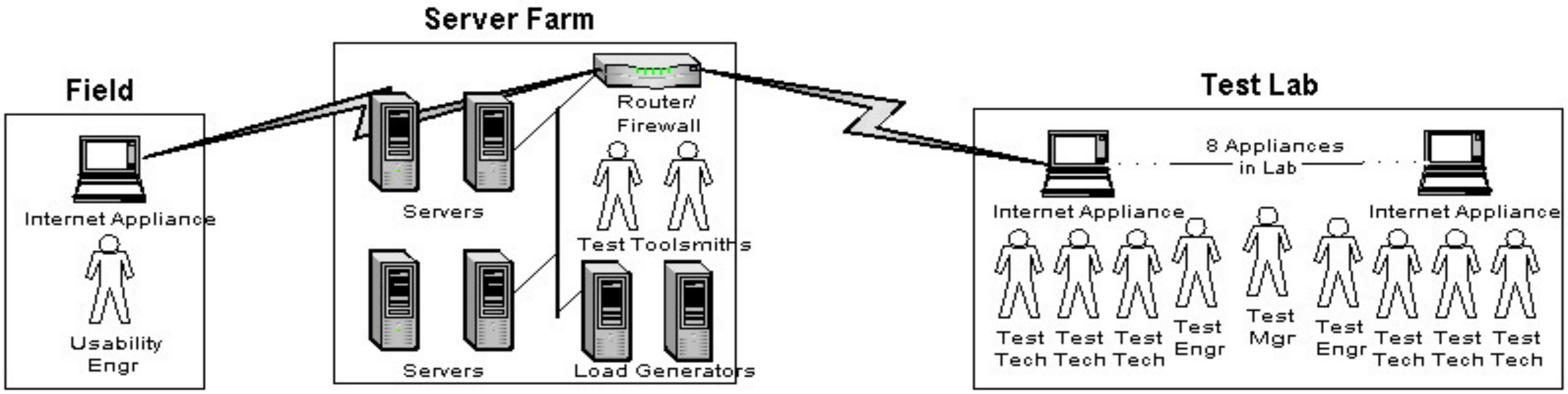
Test Technican

Test Technican

Test Technician

Test Technican

Test Technican

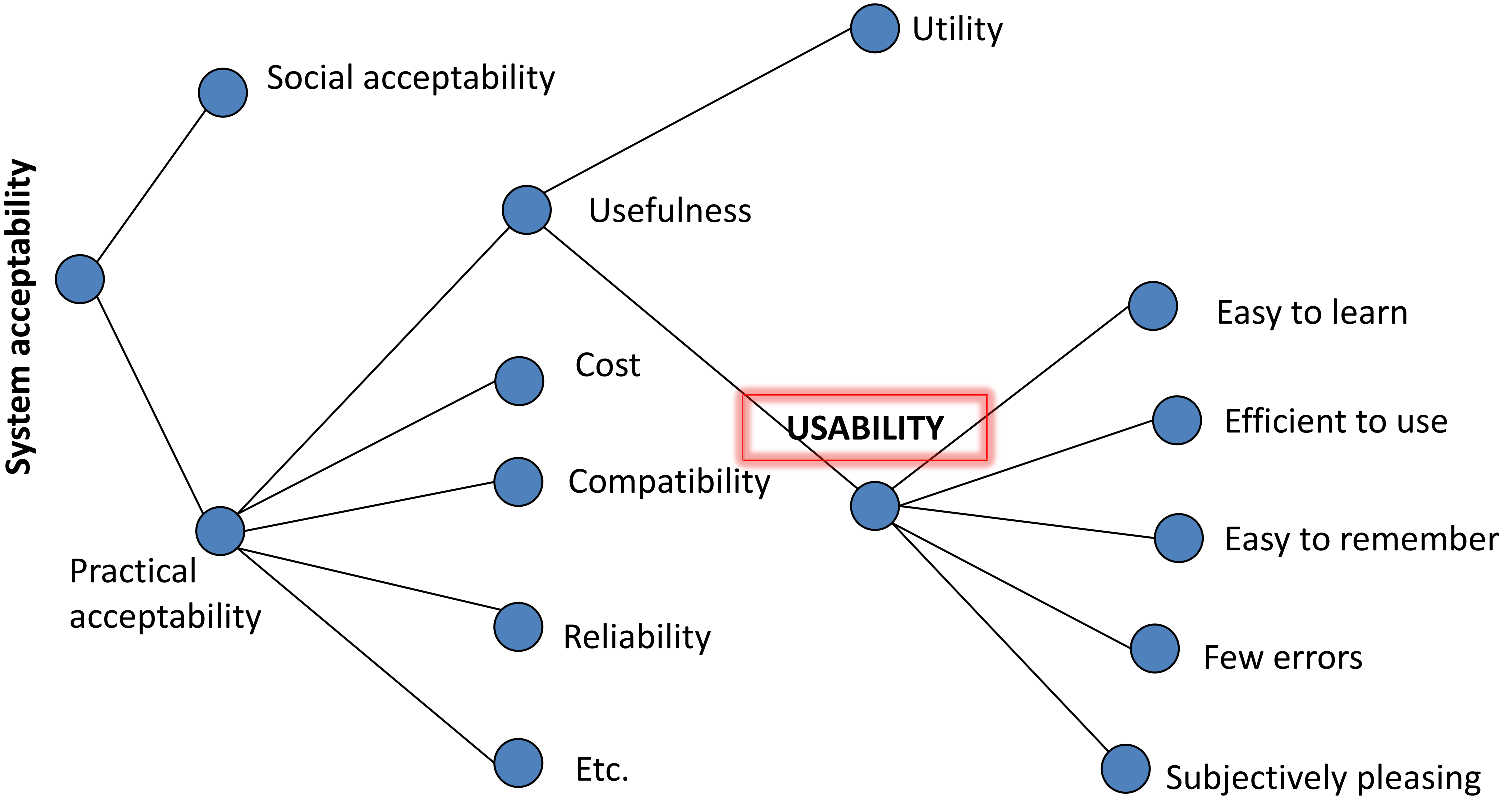


1. Use, Usability and User Experience

- Different ways to consider and evaluate the way in which we use and interact with technology
- **Use** - a general view of the qualities of digital artefacts – what products are and what they mean - but also a traditional view of technology
- **Usability** is often seen as primarily concerned with the ease of use of the user interface
- **User Experience** (**UX**) is about how people feel about a product and their pleasure and satisfaction when using it, looking at it, holding it, and opening or closing it

- We have worked mainly with usability testing, but are now looking for ways of measuring UX

Usability as an attribute of system acceptability



Nielsen, Usability engineering, p. 25

What is usability?

- Not a single, one-dimensional property of a user interface
- In the past “User friendly” – term no longer seen as appropriate
- Part of system acceptability
  - Social acceptability
  - Practical acceptability
    - Cost, Compatibility, Reliability…
    - Usefulness – if a system can be used to achieve a goal
      - Utility (Functionality – can the system do what is needed?)
      - **Usability (How well can users use that functionality?)**
        - » Easy to learn
        - » Efficient to use
        - » Easy to remember
        - » Few errors
        - » Subjectively pleasing

Nielsen, Usability engineering, p. 24

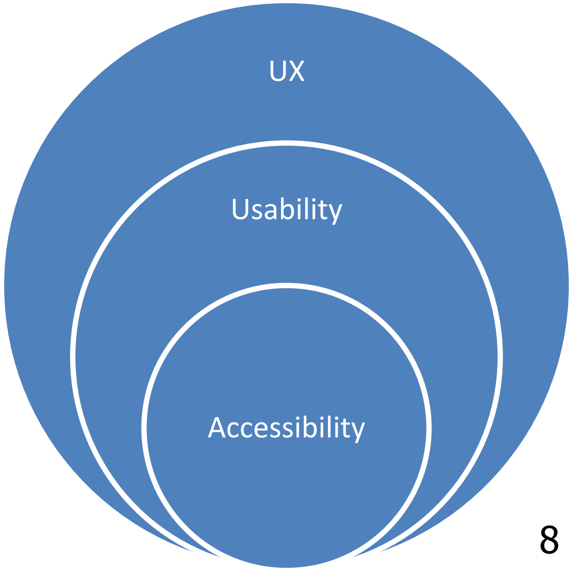
Overview on Usability

- Usability: important issue in software development
- Being concerned by an increased number of
  - Human factors specialists
  - UI designers
  - Product designers
  - Programmers
- Lack of usability
  - Annoying users, risking lives

What is usability?

1. Usability means focusing on users
2. People use products to be productive (*and for many other reasons*)
3. Users are busy people trying to accomplish tasks
4. Users decide when a product is easy to use

User Experience (UX)



What is usability?

- In the past “**User friendly**” – no longer used
- Part of system acceptability
  - Social acceptability
  - Practical acceptability
    - Cost, Compatibility, Reliability…
    - Usefulness – if a system can be used to achieve a goal
      - Utility (Functionality – can the system do what is needed?)
      - **Usability (How well can users use that functionality?)**
        - » Easy to learn
        - » Efficient to use
        - » Easy to remember
        - » Few errors
        - » Subjectively pleasing

Nielsen, Usability engineering, p. 24

Easy to learn

- Must be learnable (might need to be easy to learn)
- Learning curves
  - Novice user, high learnability = steep increase in learning at the beginning
  - Expert user, can take longer to achieve learning, then increase steeply
- “Walk up and use” systems must have zero learning time
- Initial ease of learning easy to measure
  - Pick users who have not used system, measure how long it takes to reach a specified level of proficiency, e.g. to be able to perform a certain task successfully.

Nielsen, Usability engineering, p. 27

Efficient to use

- The **“expert user’s”** level of performance
- Expertise can be judged by many criteria:
  - length of experience with a system,
  - number of hours using a system
- Typical way to measure efficiency:
  - get a representative sample of users with certain expertise, and measure how long it takes to perform specific tasks.

Nielsen, Usability engineering, p. 30

Easy to remember

- Intermittent use of a system, casual users
- Learnt how to use a system, must remember how to use it – typical for utility programs, used under exceptional circumstances, not part of everyday use
- Can be measured by
  - testing with users who have not used the system for a certain time, see how long it takes to perform a task
  - let users perform a task, use a memory test, see which details they remember or can explain, number of correct answers is score

Nielsen, Usability engineering, p. 31

Few errors

- Users should make as few errors as possible.
    - Glossary: Error = “A human action that produces an incorrect result”
- Measure errors: Count the number of actions/mistakes made by users while performing specified tasks
- We want few and non-catastrophic errors
- Varying impact of errors
    - Some are corrected straight away by the users, slow down use. The affect **efficiency** of use: time
    - Others are more catastrophic

Nielsen, Usability engineering, p. 32

Satisfaction

•Satisfaction is:

- The level of comfort users feel when using a product, how acceptable the product is as a means of achieving their goals
- More subjective than efficiency or effectiveness
- Can be difficult to measure
- May be the most important aspect of usability for products whose use is voluntary

Jordan An introduction to Usability

Subjectively pleasing

- How pleasant it is to use the system. Important for systems we do not have to use: non-work.
  - Entertainment may be more important than speed
- Objective measures of satisfaction:
  - Need experimental environments, create stress, are hard to interpret…
- Subjective measures – subjective opinion.
  - Replies from many users (subjective) are averaged giving an objective measure of pleasantness
- Short questionnaire after a de-briefing session after testing
  - Likert scale: 1 = strongly disagree … 5 = strongly agree
  - Semantic differential scale: Pleasing – Irritating; Safe – Unsafe …

Nielsen, Usability engineering, p. 33

2. Usability and Usability Testing

- Usability
  - The capability of the software to be understood, learned, used and attractive to the user when used under specified conditions
- Usability testing
  - Testing to determine the extent to which the software product is understood, easy to learn, easy to operate and attractive to the users under specified conditions

2. Characteristics of Usability Testing

- Primary goal: improving product usability
- Participants representing real users
- Participants do real tasks
- You observe what participants do and say
- You analyze the data, diagnose the problems, and recommend the changes

2. Important Issues to Determine

- The users of the system
- Goals of the users
- Environments of the use
- Measures/Criteria of effectiveness, efficiency and satisfaction (according to international standards)
  - The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use

Terms Involved

- Effectiveness: the accuracy and completeness with which users achieve specified goals
- Efficiency: the resources expended in relation to the accuracy and completeness with which users achieve goals
- Satisfaction: freedom from discomfort, and positive attitude to the use of the product
- Context of use: characteristics of the users, tasks and the organizational, and physical environments
- Goal: intended outcome
- Task: activities required to achieve a goal

2.1 Usability Goals

- Effective to use (effectiveness)
- Efficient to use (efficiency)
- Safe to use (safety)
- Having good utility (utility)
- Easy to learn (learnability)
- Easy to remember how to use (memorability)

- Goals can be turned into usability criteria

2.1.1 Setting Usability Goals

- Not all aspects can be given equal weight
- Make priorities based on analysis of users and tasks
- Discuss metrics of interest, and specify goals in terms of measures
- Goals can be set on different levels of performance
  - Specify the minimum acceptable level before release
  - Specify planned level aimed for

2.1.2 Participatory Design

- Important to have access to the people who will be using the system
- User raise questions that the developers have not even thought about asking
- Have regular meetings with users
- Use concrete and visible designs (prototypes, mock-ups, screen designs)

2.1.3 Prototyping

- Early testing can be based on prototypes
- Developed quickly and cheaply, easy to change
- Lead to a better understanding of the design
- Vertical prototype - few features, in-depth functionality
- Horizontal prototype - surface layer, whole system, no functionality
- Scenarios (use cases)

2.1.4.1 Testing Learnability

- Learning curves
  - Novice user, high learnability = steep increase in learning at the beginning
  - Expert user, can take longer to achieve learning, then increase steeply
- "Walk up and use" systems must have zero learning time
- Initial ease of learning, easy to measure
  - Pick users who have not used system, measure how long it takes to reach a specified level of proficiency, e.g. to be able to perform a certain task successfully

2.1.4.2 Testing Efficiency

- The expert user’s level of performance
- Expertise can be judged by many criteria
  - Length of experience with a system
  - Number of hours using a system
- Typical way to measure efficiency
  - Get a representative sample of users with certain expertise, and measure how long it takes to perform specific tasks

2.1.4.3Testing Memorability

- Intermittent use of a system, casual users
- Learnt how to use a system, must remember how to use it - typical for utility programs, used under exceptional circumstances, not part of everyday use
- Can be measured by
  - Testing with users who have not used the system for a certain time, see how long it takes to perform a task
  - Let users perform a task, use a memory test, see which details they remember or can explain, number of correct answers is score

2.1.4.4Testing Errors Made by Users

- Users should make as few errors as possible
- Measure errors: count the number of actions/mistakes made by users while performing specified tasks
- We want few and non-catastrophic errors
- Varying impact of errors
  - Some are corrected straight away by the users, slow down use. The affect efficiency of use: time
  - Others are more catastrophic

2.1.4.5 Testing Satisfaction

- Satisfaction is:
  - The level of comfort users feel when using a product, how acceptable the product is as a means of achieving their goals
  - More subjective than efficiency or effectiveness
  - Can be difficult to measure
  - May be the most important aspect of usability for products whose use is voluntary

Discussion

Discuss metrics of interest, and specify **usability goals** in terms of measures for a selected scenario (use case) of any software product.

- Effective to use (effectiveness)
- Efficient to use (efficiency)
- Safe to use (safety)
- Having good utility (utility)
- Easy to learn (learnability)
- Easy to remember how to use (memorability)

Goals can be turned into usability criteria

Literature

- Dumas, J.S. and Redish, J.C., 1999. A practical Guide to Usability Testing, Intellect
- Earthy, J., Jones, B. S., & Bevan, N. (2001). The improvement of human-centred processes—facing the challenge and reaping the benefit of ISO 13407. *International Journal of Human-Computer Studies*, *55*(4), 553-585.
- ISTQB: Certified tester: Foundation level syllabus (pdf and epub)
- ISTQB: Glossary
- Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003, August). The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American conference on Human-computer interaction* (pp. 53-60). ACM.
- Jordan, P.W., 2002. *An introduction to usability*. Taylor & Francis.
- Nielsen, J., 1993. Usability Engineering. Morgan Kaufmann
- Rogers, Y., Sharp, H., & Preece, J. (2011). Interaction Design: Beyond Human-Computer Interaction.
- Roto, V., Law, E., Vermeeren, A. P. O. S., & Hoonhout, J. (2011). User experience white paper. *Bringing clarity to the concept of user experience*.

Usability Testing Part II

1.1 Usability - ISTQB

- Usability is the extent to which a software product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [ISO 9241-210]. Usability testers should be aware that other definitions may be used in organizations (see [ISTQB_GLOSSARY]).
- The user interface consists of all components of a software product that provide information and controls for the user to accomplish specific tasks with the system.
- **Usability evaluation includes the following principal activities:**
  - **Usability reviews**
  - **Usability testing**
  - **User surveys**

1.2 Evaluation Types - ISTQB

Usability Evaluation

A process through which information about the usability of a system is gathered in order to improve the system (known as formative evaluation) or to assess the merit or worth of a system (known as summative evaluation).

User Experience Evaluation

User experience describes a person's perceptions and responses resulting from the use or anticipated use of a software product.

Accessibility Evaluation

Accessibility evaluation is a usability evaluation which focuses on the accessibility of a software product. It addresses the direct interaction between a user with disabilities or limitations and the software product.

1.3 Usability Evaluation - ISTQB

Usability evaluation addresses the following:

Effectiveness:

- The extent to which correct and complete goals are achieved
  Answers the question: “Does the software product do what I want?”

Efficiency:

- Resources expended to achieve specified goals
  Answers the question: “Does the software product solve my tasks quickly?”

Satisfaction:

- Freedom from discomfort, and positive attitudes towards the use of the software product
  Answers the question: “Do I feel comfortable while using the software product?”

1.4 Usability Problem Examples - ISTQB

A software product can work exactly to specification and still have serious usability problems, as shown by the following examples:

- A car rental mobile app has a dead link. This is a defect which results in a usability problem.
- A car rental mobile app allows users to cancel a reservation, but the users perceive the cancellation procedure as unreasonably complicated. This is a usability problem which affects the efficiency of the mobile app.
- A car rental mobile app conforms to the specification and works both effectively and efficiently, but users think it looks unprofessional. This is a usability problem which affects user satisfaction when using the mobile app.

Usability always relates to **the context of use** and can be considered in different components. As the following examples show, user expectations of usability are rather different for these components.

- Administrative staff use Microsoft Word ® to write documents in a consultancy firm
- An elderly person uses Microsoft Word® for the first time to write an invitation to her birthday

2. Purpose of Evaluation

- The purpose decides the methodology, data gathering, reporting
- Purpose can depend on the stage of the product creation process
- The following can be different parts of the process:
  - Benchmark existing products
  - Requirements capture
  - Evaluating initial concepts
  - Developing a concept
  - Testing a design
  - Evaluating a finished product

3. Selecting Participants

- Also depends on the purpose of the testing
- When the purpose is fixed, we must decide who will take part in the test
  - No participants (non-empirical methods)
  - Colleagues
  - A representative sample: characteristics mirror the intended users
  - Real end-users
  - The entire user population

4. Type of Data Required

- Depends on the purpose of the test. When we know the requirements, and the participants, we can consider the type of data required – dependent on the purpose, circumstances and conditions
- Quantitative
  - Performance data
  - Attitude data
- Qualitative
  - Performance data
  - Attitude data

4.1 Quantitative Data

- Performance data: time on task, error rate. Quantify effectiveness and efficiency.
  - Useful where design decisions are to be made, and there are several possibilities
  - Associated with professional products, where effectiveness and efficiency are key issues
- Attitude data: if goals are to be set for attitude, it must be possible to quantify it. It can be judged in relation to a set of criteria.
  - For example, a certain percentage of users rated the product at a certain level of the System Usability Scale.
  - Concrete criterion that can be included in a product specification.
  - Raises the status of user attitude in the same way as technical and performance aspects of a product.

4.2 Qualitative Data

- Rich descriptive data, useful for diagnosing usability faults and finding design solutions
- Performance data: Identify general areas or aspects of a product that may contain usability faults. Can be tested with quantitative methods.
  - Useful for diagnosing usability problems and finding solutions – descriptions of problems and thoughts
- Attitude data: Parallel to the qualitative performance data. Direct attention to things that are causing dissatisfaction.
  - Can provide both diagnostic and prescriptive information

5.1 Choosing Testers

- Testers (test users, participants…) should be as representative as possible of intended users of the system
- With few testers, involve average users
- With more testers, choose users from several subpopulations to cover different categories
  - The individuals who will use the system
  - Certain types of users (professions)
  - The general population

5.2 Choosing Test Leaders

- Use good test leaders with experience of the chosen method
- Should have knowledge of the application and interface:
  - to understand what the user is doing as they perform tests,
  - to make inferences about users’ intentions
- Reasons for letting the system’s designers run the tests
  - They have system knowledge
  - Seeing users’ problems has a powerful impact on them
  - BUT: they may lack objectivity. They might explain away problems rather than accept them as problems

5.3 Choosing Tasks

- As representative as possible of real-life use
- Cover the most important parts of the system
- Small enough to be completed within a reasonable time, but not trivial
- Specify what the user is expected to do
- Give users tasks in writing
- Allow users to ask questions about the task description
- Start with a simple task
- The final task should let the user feel that they have accomplished something

6. Usability Evaluation Approaches

- Non-empirical
  - No participant required
- Empirical
  - Real users use the product

6.1 Non-Empirical Approaches

- Task analysis
- Property checklist (heuristics evaluation)
- Expert appraisals
- Cognitive walkthroughs
  - The expert approaches the task from the point of view of a typical user, with a typical task

6.2 Non-Empirical Methods (cont.)

- Heuristic evaluation
  - Look at an interface: what is good and bad
  - Guidelines, often based on intuition and common sense
  - Each evaluator inspects the interface alone, and then the evaluators communicate, and results are aggregated - independent & unbiased
  - Can be performed on early prototypes
  - Output is a list of usability problems, with references to violated principles
- Cognitive walkthrough
  - Evaluate the strengths and weaknesses of an interface
  - Deals with users' goals, and how easily they can progress towards their goals, for each step of the process
  - Depends on a thorough understanding of the users and their skills

6.2 How should we evaluate usability?

	# of problems	% of problems
Usability test	25	100%
Heuristic evaluation		
Experts	11	44
Software engineers	4	16%
Non-experts	2	8%
Cognitive walkthrough		
Experts	7	28%
Software engineers	4	16%
Non-experts	2	8%

Dumas, Redish, A practical Guide to Usability Testing. P. 81

6.2 What to choose?

- Perform **both heuristic evaluations and usability tests** – take advantage of the strengths of both
- The best evaluation comes from **multiple techniques**
- Usability tests uncover global problems, and problems other methods miss
- If you only do usability tests, you risk missing local problems, that can irritate and frustrate users – local problems can make users think the product design is sloppy

Dumas, Redish, A practical Guide to Usability Testing. P. 82

6.2 Empirical Methods

- Private camera conversations
- Co-discovery: two participants working together to explore a product and/or test how tasks are done
- Focus groups - people gathered together to discuss a particular issue
- User workshops - a group of people gathered together to discuss issues connected to design
- Think-aloud protocols - a participant speaking about what they are doing or thinking when using an interface
- Incident diaries - making note of problems when using a product
- Feature checklists - a list of functions, the user marks those used

6.2 Empirical Methods (cont.)

- Logging use - automatic logging that records interaction with product
- Field observation - watching users in the environment where they would normally use a product
- Questionnaires - fixed-response or open-response
- Interviews - a series of questions that are posed directly to participants: unstructured, semi-structured and structured
- Valuation methods - evaluate comparative importance by asking how much extra they would pay for a certain feature
- Controlled experiments - a formally designed evaluation with relatively tight controls

Usability Testing Part III

Design and Conduct Usability Testing

Assignment 2.2: Usability Evaluation

- Each team is required to design and plan a usability evaluation on a selected scenario (use case) of any hotel booking website or flight booking website. Present the selected website and the selected scenario (use case) first, and then describe the following elements of the usability evaluation:

1) Purpose/Primary Goals – Criteria - Questions
2) Constraints and opportunities
3) Questions for the evaluation
4) Approach to answering questions (Non-Empirical/Empirical)
5) Data needed (Criteria-Question- Metrics)
6) Methods to be used (how to collected the data based on metrics)
7) How to report the results (analysis results, report, and improve recommendations)

Testing – scenarios (use cases)

- They tell participants what you want them to do.
- They describe tasks in a way that removes artificiality
- A good scenario:
  - Is short
  - Is in the user's words, not the product's
  - Is clear, so participants understand it
  - Gives enough information to do the task
  - Is directly linked to your tasks and concerns
- A scenario can be given verbally or (more often) in writing
- A scenario can contain several subtasks, or separate tasks

Dumas, Redish, A practical Guide to Usability Testing. P. 89

1. Defining Goals and Concerns

- Decide **what to focus upon** – you can't test everything with everyone
- Deciding **goals and concerns affects all other decisions** – who should participate, what tasks to test, what equipment is needed, what to measure.
- You can begin by **planning goals**, or **concerns**, or a **combination** of both
- A **goal** is usually a **statement**:
  - "Users will be able to select the correct icon in less than 30 seconds, with no more than one mistake"
- A **concern** is usually in the form of a **question**:
  - "Will users be able to correct the right icon quickly and accurately?"

Dumas, Redish, A practical Guide to Usability Testing. P. 110

1.1 Usability goals

- Primary goal: improving product usability
  - Participants representing real users
  - Participants do real tasks
  - You observe what participants do and say
  - You analyze the data, diagnose the problems, and recommend the changes
- Sub- Goals
  - Effective to use (effectiveness)
  - Efficient to use (efficiency)
  - Safe to use (safety)
  - Having good utility (utility)
  - Easy to learn (learnability)
  - Easy to remember how to use (memorability)
- Goals can be turned into usability criteria

1.2 Goals as questions

- Usability goals can be operationalized as questions, to provide the designer with a concrete means of assessing aspects of the product and the user experience
- By answering the questions, the designer can be alerted early in the process to potential design problems and conflicts that they may not have considered

Rogers, Sharp & Preece, 2011

1.3 Quantitative usability goals

- Key component: setting **specific quantitative usability goals** for a product
- Can come from task analysis: a list of tasks that users wish to do.
- Tasks lead to usability goals. Answers the question: “How will you know if the product is easy for people to use when doing this task?”
- Can come from company goals to reduce cost and increase productivity: Reduce calls to support by 50%
- Each goal leads to design decisions to answer the question: “How are you going to make sure that the product meets this usability goal?”
- Must be a team effort: Designers, developers, marketing, interface designers…

Dumas, Redish. A practical Guide to Usability Testing. P. 48.

2. Constraints and opportunities

- Circumstances mean that there will almost certainly be constraints. There may also be opportunities that can be taken advantage of.
  - Deadlines
  - Investigator time
  - Participant time
  - Investigator knowledge
  - Participants available
  - Facilities and resources

Jordan An introduction to Usability

3. What Can You Measure

- Performance measures: counts of actions and behaviors that you can see
  - Quantitative: how much time someone needs, how many errors they make, how many times they repeat the errors
  - Require observation, but usually not judgement (unless you are counting e.g. expressions of discontent, irritation, joy…)
- Time: finishing a task; navigating menus; finding information in help menus; reading manuals; recovering from errors
- Number: of clicks to complete a task; of wrong menu choices; of incorrect choices in dialogues; of other errors; of repeated errors; of screens of online help looked at; of looks at same help screen; of times using manuals;…
- Observations: of frustration; of confusion; of satisfaction…

3. What Can You Measure (cont.)

- Subjective measures: people’s perceptions, opinions and judgements
- Quantitative or qualitative. Likert scales give a subjective judgement, but the response is quantitative. Spontaneous comments when thinking aloud are qualitative, but you can report frequencies – how many comments on a particular issue
- Ratings of: ease of learning; ease of use; ease of doing a certain task; helpfulness of help functions; ease of understanding information;…
- Preferences and reasons for preferences: over a previous version; over a competitor’s version; over design alternatives; over how they do things now;…
- Predictions of behavior and reasons for predicted behavior: would you buy this product? Would you pay extra for the manual; How much would you pay for this product?
- Spontaneous comments: I’m lost here; That was easy; I think I’d call support; I don’t understand the message;… (Think aloud)

3.1 Match measures to goals and concerns

- Measures chosen should be related to the quantitative usability goals that you set, and the concerns that drive the test
- Every goal and concern should lead to one or more performance measures
  - Concern: Right menu choices. Count: Wrong menu choices
  - Concern: Learning icons. Count: Wrong icon choices, repeated errors
- **Collect only what you need** to measure your goals and concerns
- Remember, you cannot cover everything in one test.

Dumas, Redish, A practical Guide to Usability Testing. P. 189

3.1 Matching measures to development stage

- Where is the product in the development stage?
- For example, **Testing a prototype**:
  - Performance measures like wrong menu choice, repeated errors, frustration…
  - Time is probably not important, especially if you are not on running the goal system
- Iterative testing at different stages of development:
  - Repeat some measures to check improvements
  - Collect other measures, because you will want to check different product aspects at different development stages

Dumas, Redish, A practical Guide to Usability Testing. P. 194

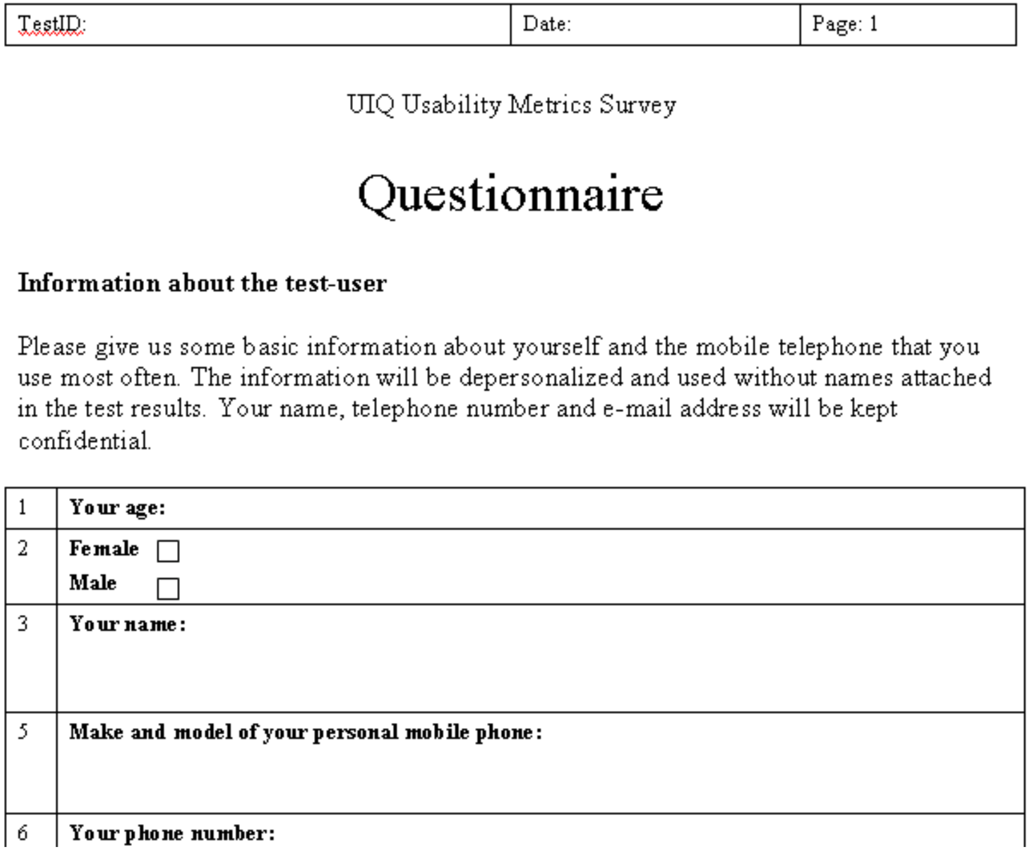
3.2 Setting Usability criteria

- Think in advance how you are going to interpret the data
- It is not enough to say a product is “usable” or “easy to use”. You must define it more explicitly and quantitatively

- Performance measures: counts of actions and behaviors that you can see
  - Quantitative: how much time someone needs, how many errors they make, how many times they repeat the errors
  - Require observation, but usually not judgement (unless you are counting e.g. expressions of discontent, irritation, joy…)
- Time: finishing a task; navigating menus; finding information in help menus; reading manuals; recovering from errors
- Number: of clicks to complete a task; of wrong menu choices; of incorrect choices in dialogues; of other errors; of repeated errors; of screens of online help looked at; of looks at same help screen; of times using manuals;…
- Observations: of frustration; of confusion; of satisfaction…

Example : The questionnaire

- The first step in the UTUM test is to fill in the questionnaire
- This is done in cooperation with the tester
- In the questionnaire we collect data that can have an effect on the result



TestID:	Date:	Page: 1

UIQ Usability Metrics Survey

Questionnaire

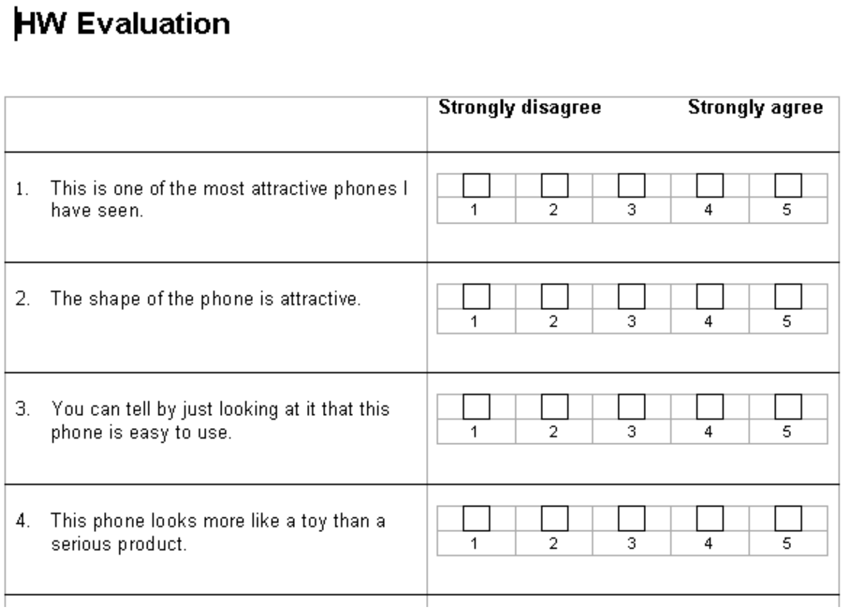
Information about the test-user

Please give us some basic information about yourself and the mobile telephone that you use most often. The information will be depersonalized and used without names attached in the test results. Your name, telephone number and e-mail address will be kept confidential.

1	**Your age:**
2	**Female** □ **Male** □
3	**Your name:**
5	**Make and model of your personal mobile phone:**
6	**Your phone number:**

Example : Hardware evaluation

- When the user has tried the phones.
- Before the use case tests
- They fill in a evaluation questionnaire regarding specific look and feel of the device.

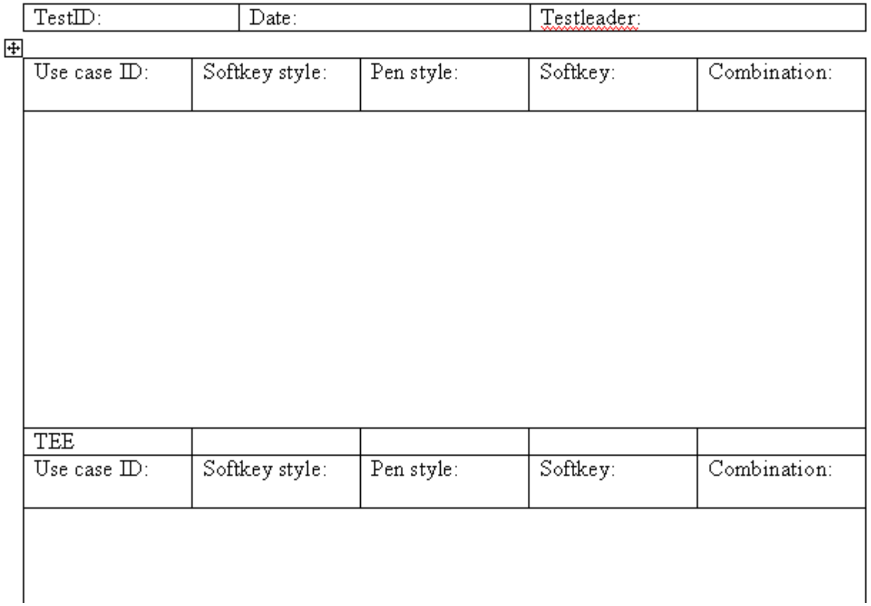


HW Evaluation

	Strongly disagree				Strongly agree
1. This is one of the most attractive phones I have seen.	□ 1	□ 2	□ 3	□ 4	□ 5
2. The shape of the phone is attractive.	□ 1	□ 2	□ 3	□ 4	□ 5
3. You can tell by just looking at it that this phone is easy to use.	□ 1	□ 2	□ 3	□ 4	□ 5
4. This phone looks more like a toy than a serious product.	□ 1	□ 2	□ 3	□ 4	□ 5

Example : Performing the use case

- While the tester performs a use case the test leader times it, writing down any setbacks.
- Observe the users experience
- Fill in the result in the test notes
- And rank the result on a scale between 0-4.

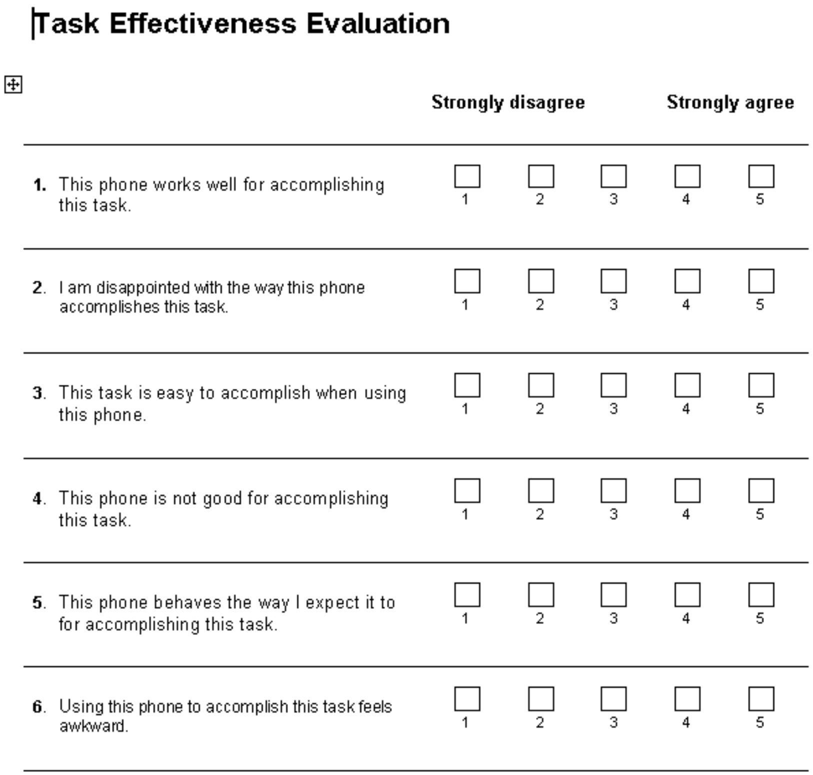


TestID:	Date:	Testleader:

Use case ID:	Softkey style:	Pen style:	Softkey:	Combination:
TEE				
Use case ID:	Softkey style:	Pen style:	Softkey:	Combination:

Example : Task Effectiveness Evaluation

- After each use case, the tester fills in a Task Effectiveness Evaluation
- Based on how they experienced the use case

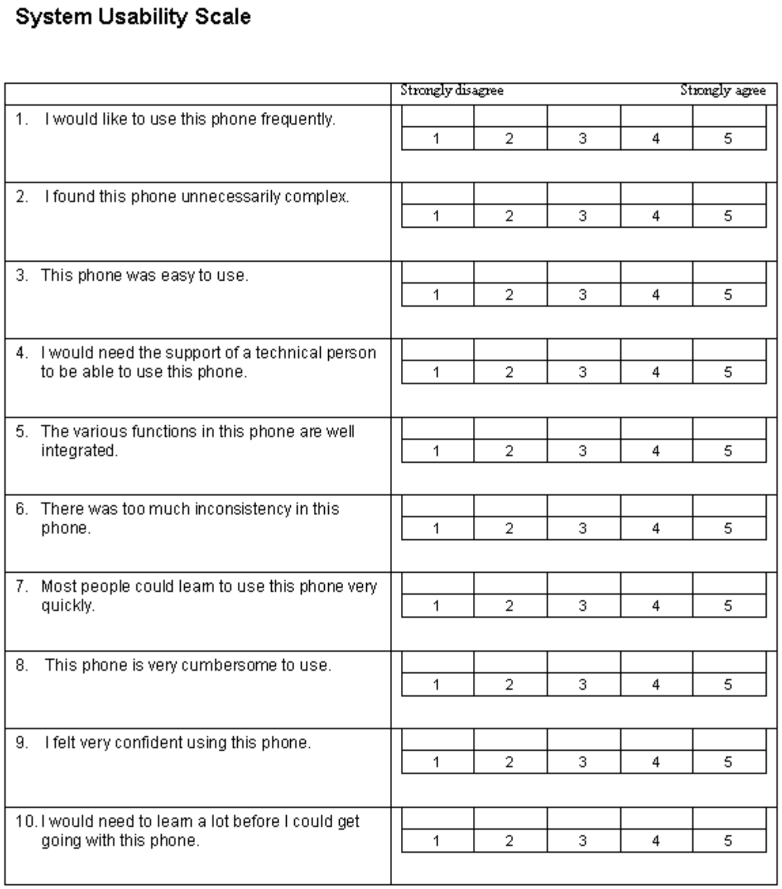


Task Effectiveness Evaluation

	Strongly disagree				Strongly agree
1. This phone works well for accomplishing this task.	□ 1	□ 2	□ 3	□ 4	□ 5
2. I am disappointed with the way this phone accomplishes this task.	□ 1	□ 2	□ 3	□ 4	□ 5
3. This task is easy to accomplish when using this phone.	□ 1	□ 2	□ 3	□ 4	□ 5
4. This phone is not good for accomplishing this task.	□ 1	□ 2	□ 3	□ 4	□ 5
5. This phone behaves the way I expect it to for accomplishing this task.	□ 1	□ 2	□ 3	□ 4	□ 5
6. Using this phone to accomplish this task feels awkward.	□ 1	□ 2	□ 3	□ 4	□ 5

Example : System Usability Scale

- After all use cases are completed the tester has to fill in a
System Usability Scale.
- This part is done interactively with the test leader.
- The test leader makes additional notes in the UseCase test note document.

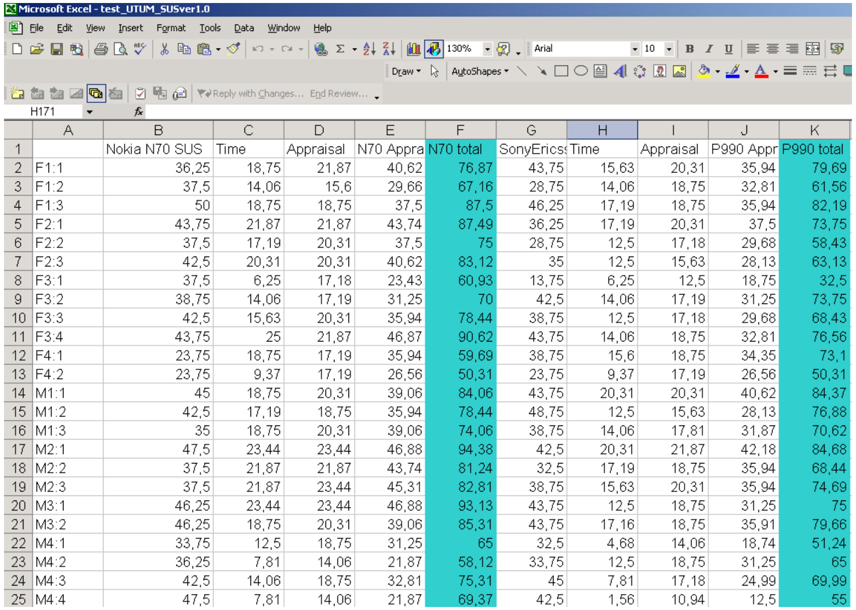


System Usability Scale

	Strongly disagree				Strongly agree
1. I would like to use this phone frequently.	1	2	3	4	5
2. I found this phone unnecessarily complex.	1	2	3	4	5
3. This phone was easy to use.	1	2	3	4	5
4. I would need the support of a technical person to be able to use this phone.	1	2	3	4	5
5. The various functions in this phone are well integrated.	1	2	3	4	5
6. There was too much inconsistency in this phone.	1	2	3	4	5
7. Most people could learn to use this phone very quickly.	1	2	3	4	5
8. This phone is very cumbersome to use.	1	2	3	4	5
9. I felt very confident using this phone.	1	2	3	4	5
10. I would need to learn a lot before I could get going with this phone.	1	2	3	4	5

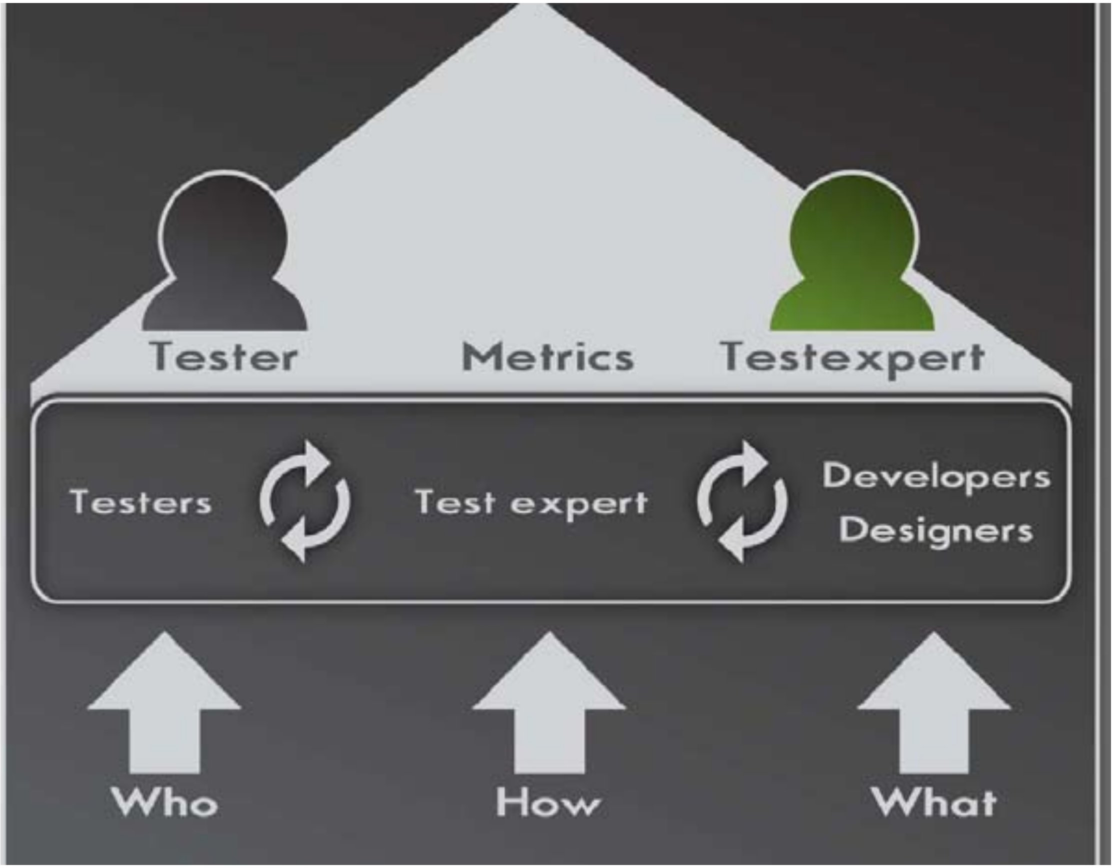
Example : Saving the data

- Test data is collected in a spreadsheet for further analysis


Microsoft Excel - test_UTUM_SUSver1.0
File Edit View Insert Format Tools Data Window Help
130%
Arial
10
Draw
AutoShapes
Reply with Changes... End Review...
H171
fx

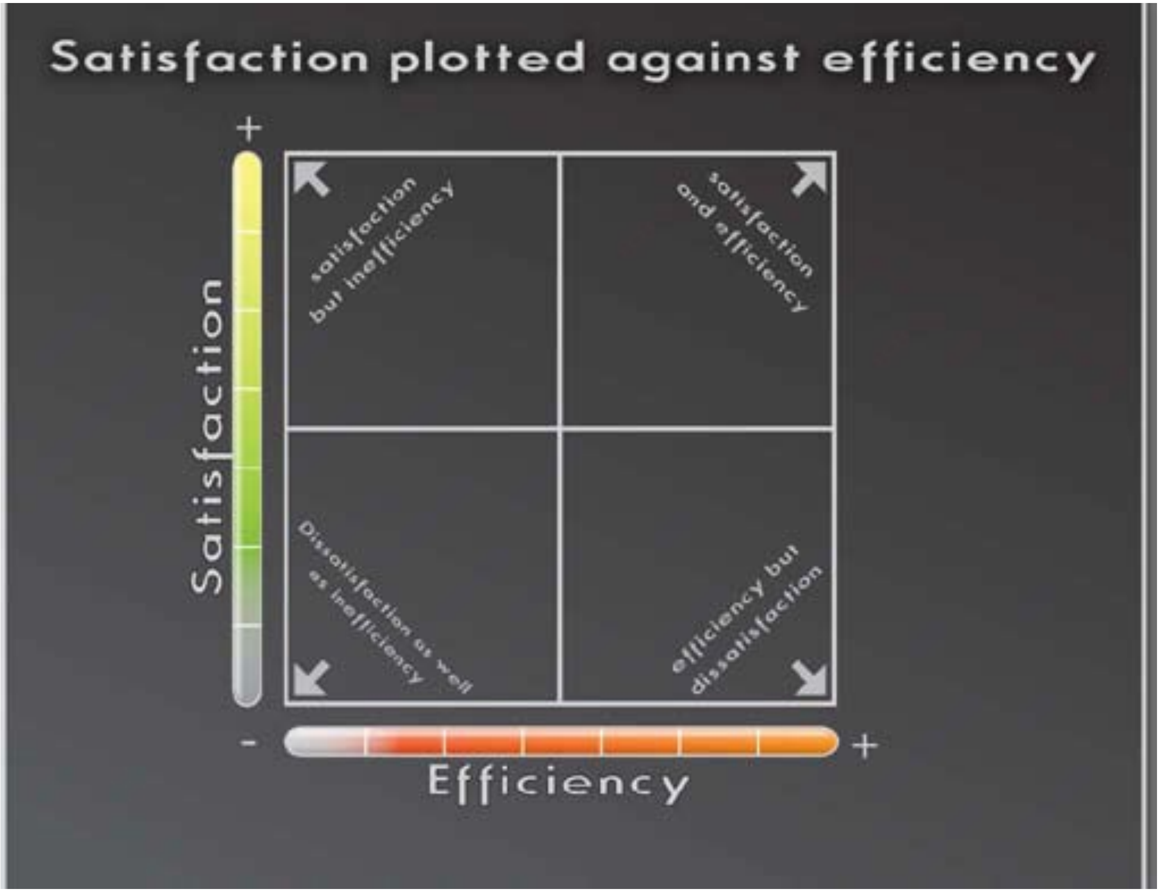
	A	B	C	D	E	F	G	H	I	J	K
1		Nokia N70 SUS	Time	Appraisal	N70 Appra	N70 total	SonyEricss	Time	Appraisal	P990 Appr	P990 total
2	F1:1	36,25	18,75	21,87	40,62	76,87	43,75	15,63	20,31	35,94	79,69
3	F1:2	37,5	14,06	15,6	29,66	67,16	28,75	14,06	18,75	32,81	61,56
4	F1:3	50	18,75	18,75	37,5	87,5	46,25	17,19	18,75	35,94	82,19
5	F2:1	43,75	21,87	21,87	43,74	87,49	36,25	17,19	20,31	37,5	73,75
6	F2:2	37,5	17,19	20,31	37,5	75	28,75	12,5	17,18	29,68	58,43
7	F2:3	42,5	20,31	20,31	40,62	83,12	35	12,5	15,63	28,13	63,13
8	F3:1	37,5	6,25	17,18	23,43	60,93	13,75	6,25	12,5	18,75	32,5
9	F3:2	38,75	14,06	17,19	31,25	70	42,5	14,06	17,19	31,25	73,75
10	F3:3	42,5	15,63	20,31	35,94	78,44	38,75	12,5	17,18	29,68	68,43
11	F3:4	43,75	25	21,87	46,87	90,62	43,75	14,06	18,75	32,81	76,56
12	F4:1	23,75	18,75	17,19	35,94	59,69	38,75	15,6	18,75	34,35	73,1
13	F4:2	23,75	9,37	17,19	26,56	50,31	23,75	9,37	17,19	26,56	50,31
14	M1:1	45	18,75	20,31	39,06	84,06	43,75	20,31	20,31	40,62	84,37
15	M1:2	42,5	17,19	18,75	35,94	78,44	48,75	12,5	15,63	28,13	76,88
16	M1:3	35	18,75	20,31	39,06	74,06	38,75	14,06	17,81	31,87	70,62
17	M2:1	47,5	23,44	23,44	46,88	94,38	42,5	20,31	21,87	42,18	84,68
18	M2:2	37,5	21,87	21,87	43,74	81,24	32,5	17,19	18,75	35,94	68,44
19	M2:3	37,5	21,87	23,44	45,31	82,81	38,75	15,63	20,31	35,94	74,69
20	M3:1	46,25	23,44	23,44	46,88	93,13	43,75	12,5	18,75	31,25	75
21	M3:2	46,25	18,75	20,31	39,06	85,31	43,75	17,16	18,75	35,91	79,66
22	M4:1	33,75	12,5	18,75	31,25	65	32,5	4,68	14,06	18,74	51,24
23	M4:2	36,25	7,81	14,06	21,87	58,12	33,75	12,5	18,75	31,25	65
24	M4:3	42,5	14,06	18,75	32,81	75,31	45	7,81	17,18	24,99	69,99
25	M4:4	47,5	7,81	14,06	21,87	69,37	42,5	1,56	10,94	12,5	55

Example : Structure of the user testing



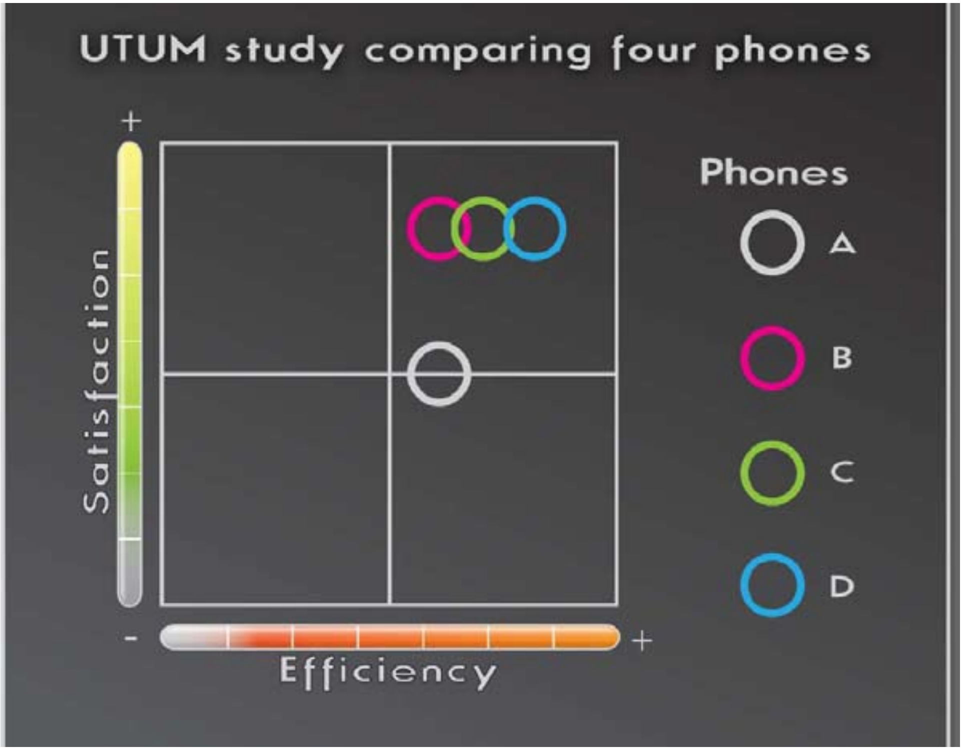
(Rönkkö, Winter, Hellman, 2009)

Example : Spatial presentation of results


Satisfaction plotted against efficiency
+
Satisfaction
satisfaction but inefficiency
satisfaction and efficiency
Dissatisfaction as well as inefficiency
efficiency but dissatisfaction
-
+
Efficiency

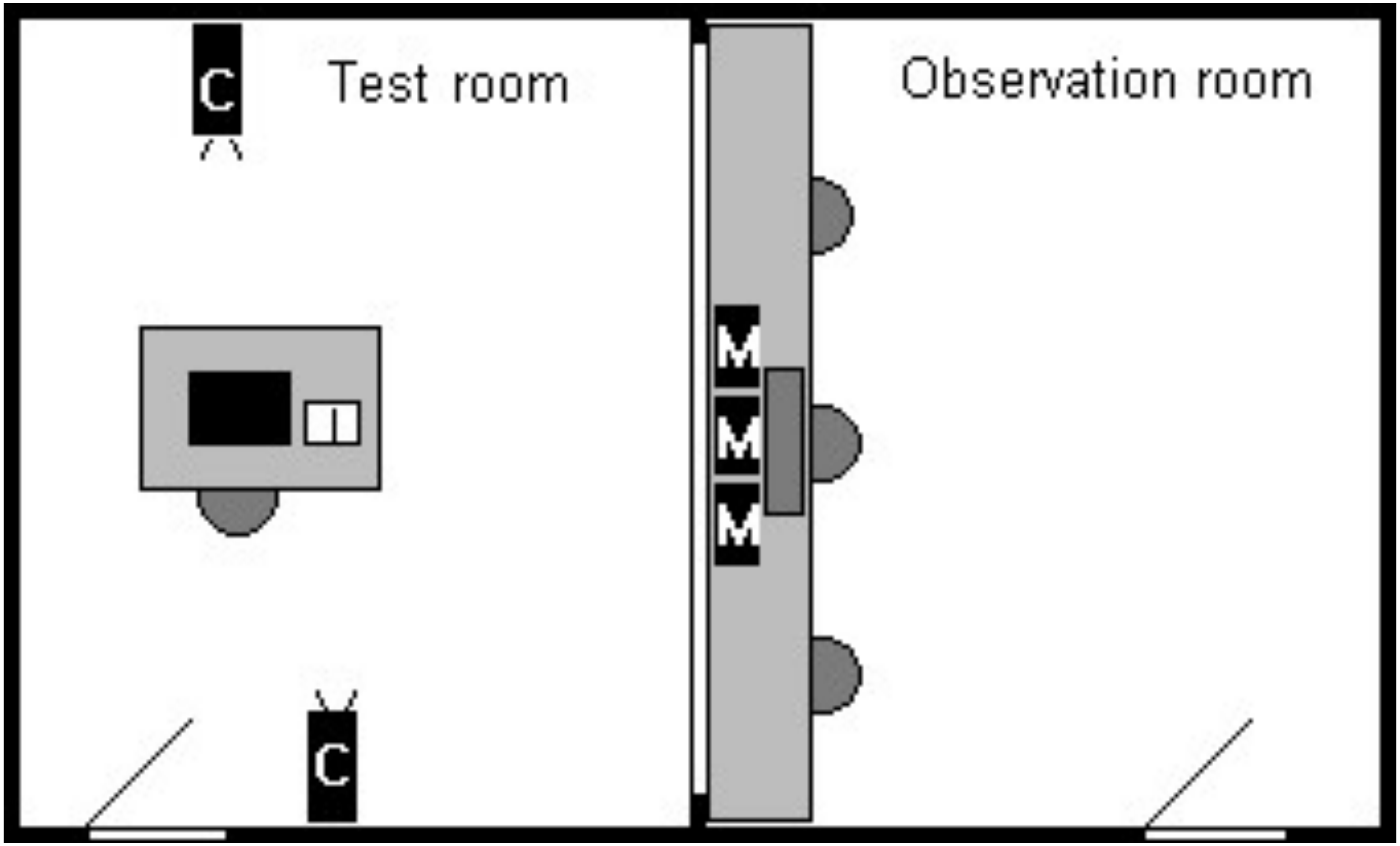
(Rönkkö, Winter, Hellman, 2009)

Example : Visualising comparisons



(Rönkkö, Winter, Hellman, 2009)

Example : Layout of Usability Test Lab – ISTQB



Layout of Usability Test Lab

4. Conduct Usability Evaluation - Stages in Testing

- A test usually has four stages
  - Preparation
  - Introduction
  - Testing
  - Debriefing

4.1 Preparation and Introduction

- Prepare everything before the user arrives
- Remind the user that participation is voluntary
- Introduce procedure and explain the purpose of the test - emphasize that the system is being tested, not the user
- Say that the results of the test are confidential, identity will not be revealed
- Explain the use of video or audio recording (if used)
- Say that the user may ask questions, since we want to know what is unclear in the interface, but that the test leader will not answer most questions during the testing
- Give any specific instructions (e.g. think out loud, try to work as fast as possible)
- Let the user ask any questions before starting

4.2 Testing

- Do not interact with the user, do not express any personal opinions – any response from the tester should be neutral
- Do not help the user
  - You may help if the user gets stuck on a problem other users have experience, and you understand the cause, and what other problems may result
- Give no help if you are timing use-case performance
- Even if there are observers, only one person is the test leader and can interact with the tester, to avoid confusing the tester

4.3 Debriefing

- Fill in any questionnaires
- Before discussing the system, to avoid bias from comments by the test leader
- Ask for comments the tester has about the system or any suggestions for improvement
- Comments may be contradictory and may not lead to changes, but can be a rich source of design ideas
- Ask users about things in the test that the test leader found hard to understand
- After the user has left, check that all of the test results have been properly recorded and connected to the tester
- Write a report of the test while the events are fresh in your mind

4.4 Reporting the evaluation

- **To those who need to know**: Important to adapt the report to the target group. Results must be **clear and persuasive**.
- Conflicting issues.
  - Engineers set technical constraints, that designers find as limiting.
  - Designers can make “imaginative” proposals that engineers see as causing unnecessary technical complications.
- Important to understand “social dynamics” between groups, and adapt presentation of results appropriately.
  - **Written reports**
  - **Verbal presentations**
  - **Summary videos**

Jordan An introduction to Usability

4.4.1 Verbal Presentations

- Formal: Slides shown to a room full of stakeholders. A detailed explanation of the evaluation, with the chance to ask questions
- Informal: Chat with others over a cup of coffee. A loose presentation of some of the findings, recommendations and their implications, to see which way the conversation develops
  - For Informal discussion, participants can prepare by reading a report or watching a presentation.
- The main advantage: the chance to ask questions, to interact with the investigator – clarifications, discussion of implications…

4.4.2 Summary Videos

- Often combined with written or verbal presentation
- Show clearly participants’ problems
- Convince others of the significance:
  - Can be hard to explain verbally or in writing, where the investigator is reporting “on trust”
  - When reporting, what should the audience make of what is descried?
  - If a problem is described as serious, what does that mean for the user?
- A video can answer these questions for the viewer

4.4.3 Testing Scenarios (Use Cases)

- They tell participants what you want them to do
- They describe tasks in a way that removes artificiality
- A good scenario:
  - Is short
  - Is in the user’s words, not the product’s
  - Is clear, so participants understand it
  - Gives enough information to do the task
  - Is directly linked to your tasks and concerns
- A scenario can be given verbally or (more often) in writing
- A scenario can contain several subtasks, or separate tasks

4.5 Written Reports

- Varying degrees of formality, depending on purpose, audience, organizational culture. A full length formal report may include:
  - Summary
  - Introduction
  - Methods
  - Results
  - Discussion
  - Conclusions
  - Recommendations

Example: Presentation methods

1. Structured Data Summary (SDS): Spreadsheet with qualitative findings of the testing. Issues for each use case, and user comments
2. Spreadsheet with all “raw data”: All the quantitative data from the testing
3. Curve Diagram: Graph showing comparison of time taken to complete one use case, showing average time for all phones, and individual plots for each phone
4. Comparison of two factors (basic): A visual comparison of three phones, for satisfaction and efficiency
5. Comparison of two factors (brief details): Same as 4, but with a brief explanation of the findings
6. Comparison of two factors (in depth details): Same as 4, but with a more extensive explanation of the results, and the findings of the test leader, plus comments for long and short term solutions to issues

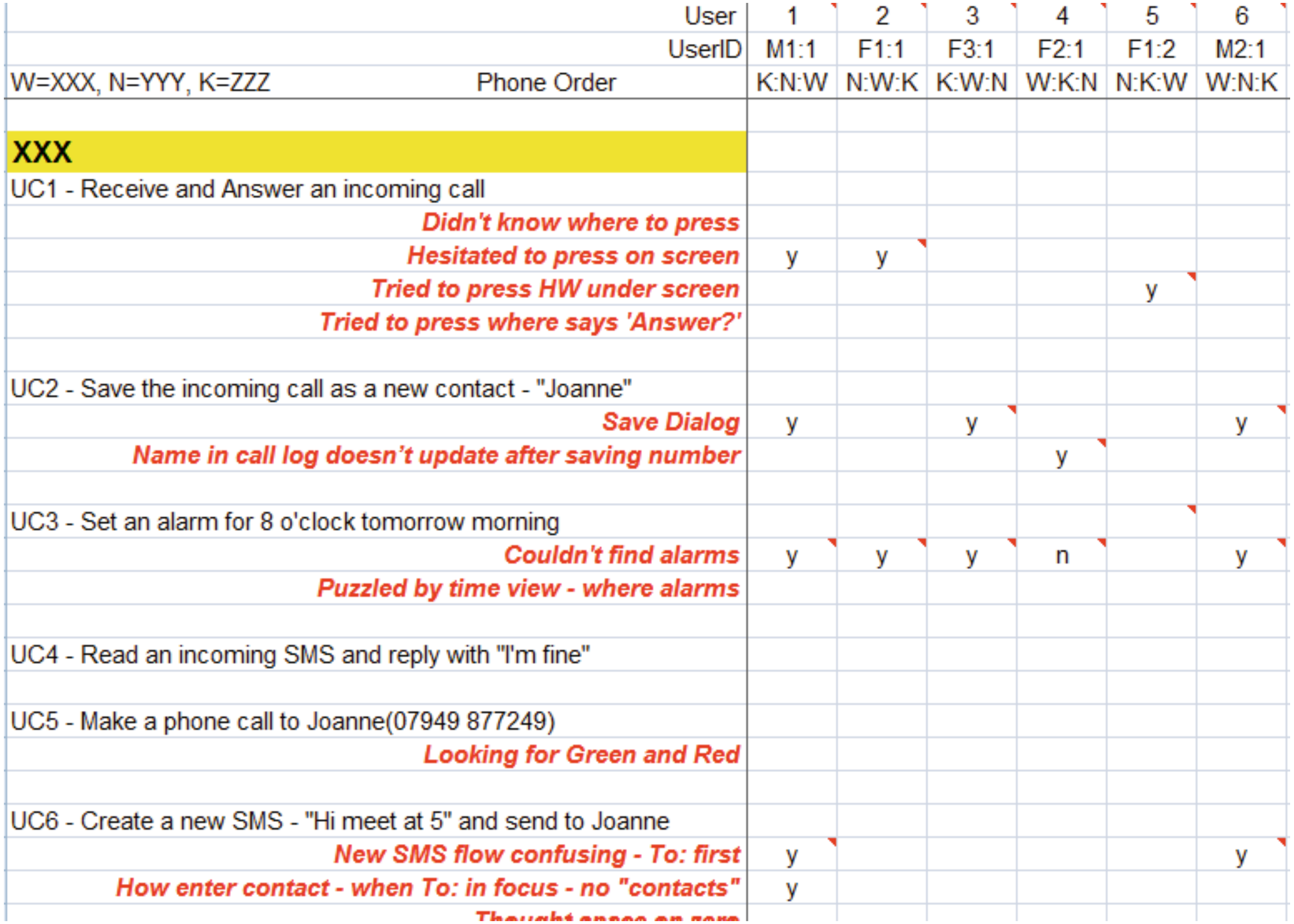
(Winter, Rönkkö, 2009)

Example: Presentation methods

7. The “Form factor”, an immediate response: A visual representation of which telephone was preferred by men and women. An immediate response to the phones, choosing a favourite phone on the basis of “Form Factor”
8. PowerPoint, no verbal presentation: Produced by the test leader. A summary of the main results is presented graphically and briefly in writing.
9. Verbal presentation with PowerPoint: Given by the test leader. Graphically and in writing, with verbal commentary
10. Verbal presentation of early results: The test leader gives a verbal presentation of the results of a series of tests, based mainly on his or her impressions of issues found, rather than an analysis of the metrics.

(Winter, Rönkkö, 2009)

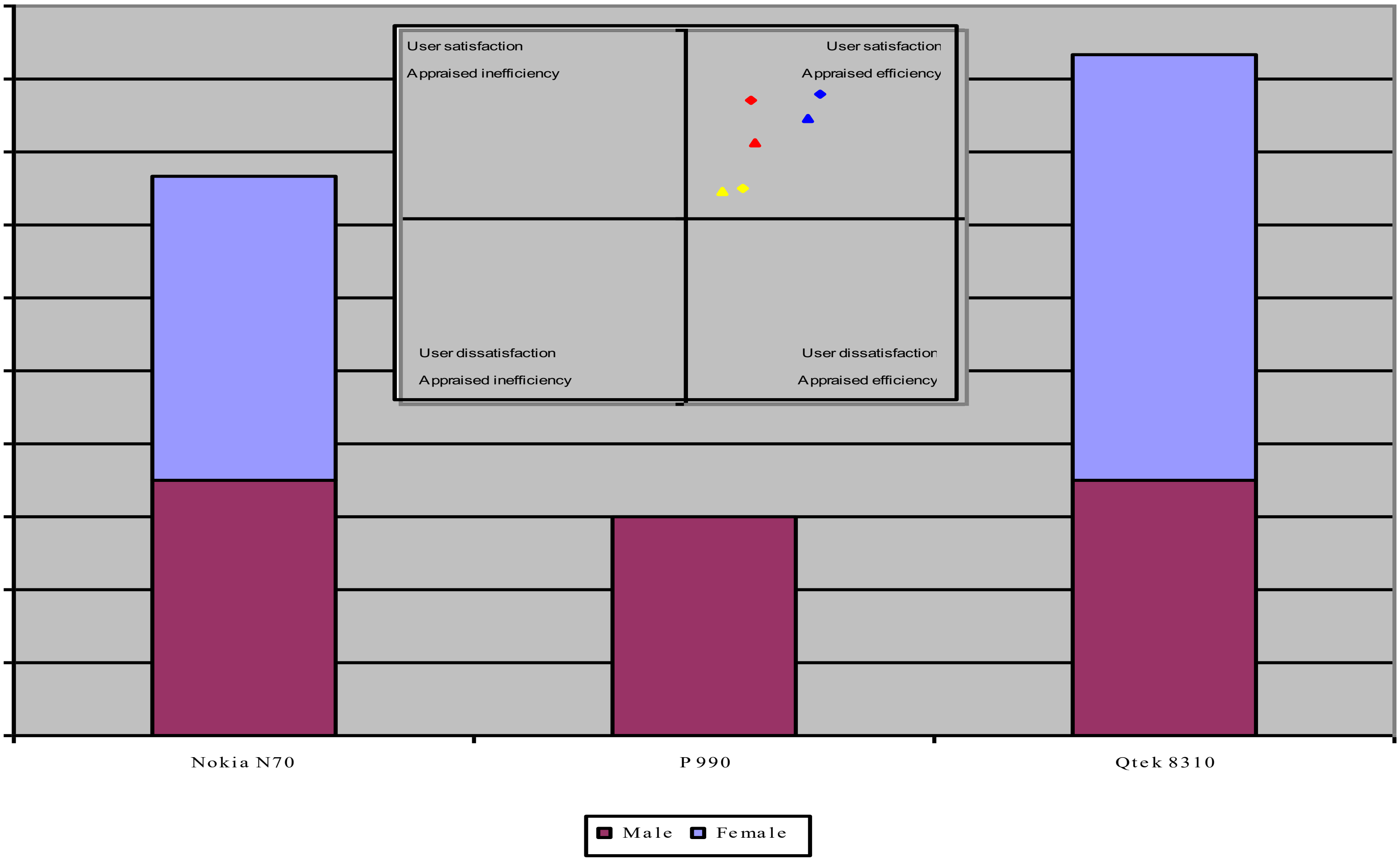
Example: Structured Data Summary



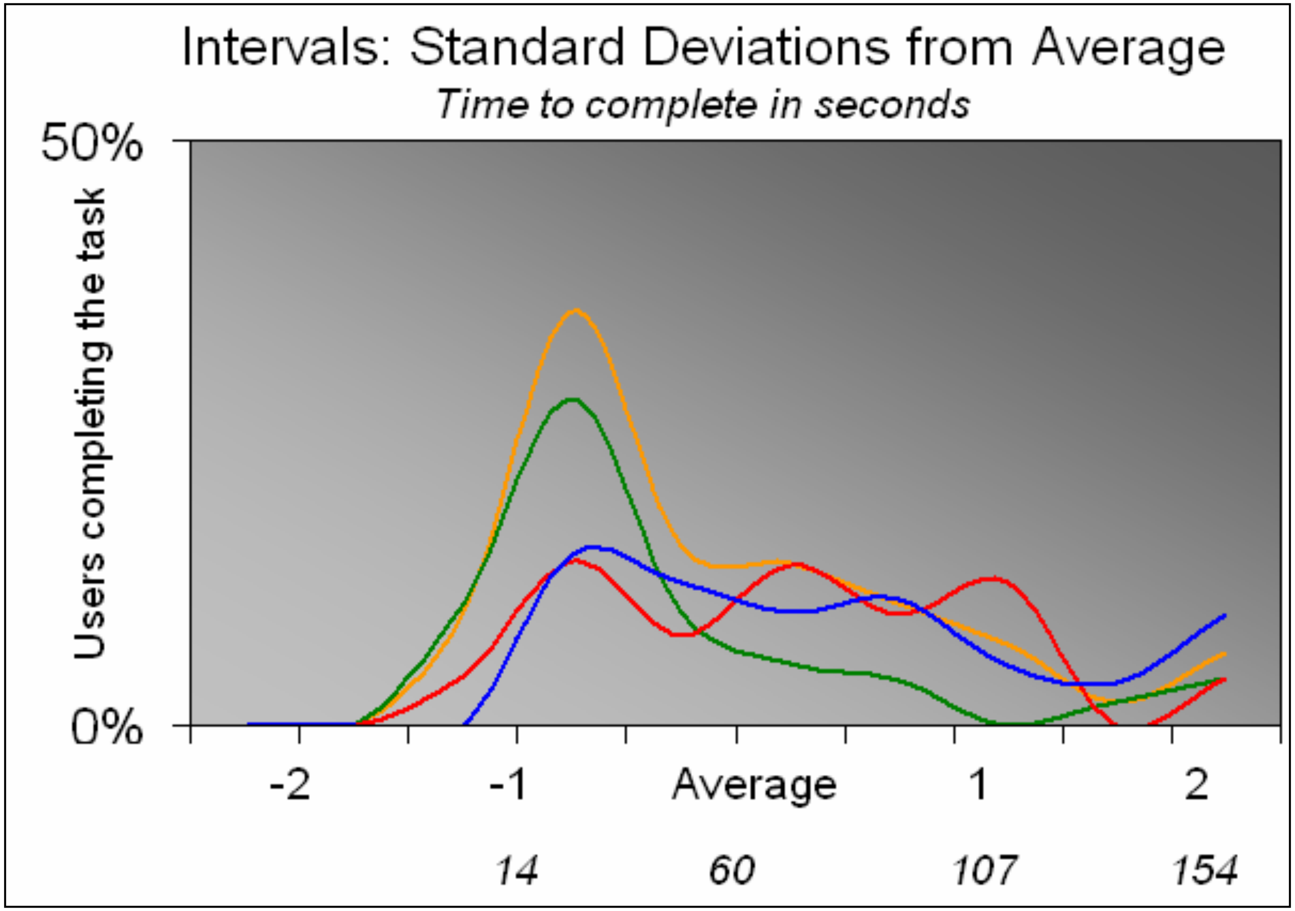
User	1	2	3	4	5	6
UserID	M1:1	F1:1	F3:1	F2:1	F1:2	M2:1
W=XXX, N=YYY, K=ZZZ Phone Order	K:N:W	N:W:K	K:W:N	W:K:N	N:K:W	W:N:K
XXX						
UC1 - Receive and Answer an incoming call						
Didn't know where to press						
Hesitated to press on screen	y	y				
Tried to press HW under screen					y	
Tried to press where says 'Answer?'						
UC2 - Save the incoming call as a new contact - "Joanne"						
Save Dialog	y		y			y
Name in call log doesn't update after saving number				y		
UC3 - Set an alarm for 8 o'clock tomorrow morning						
Couldn't find alarms	y	y	y	n		y
Puzzled by time view - where alarms						
UC4 - Read an incoming SMS and reply with "I'm fine"						
UC5 - Make a phone call to Joanne(07949 877249)						
Looking for Green and Red						
UC6 - Create a new SMS - "Hi meet at 5" and send to Joanne						
New SMS flow confusing - To: first	y					y
How enter contact - when To: in focus - no "contacts"	y					
Thought space on zero						

(Winter et al 2008 CEE SETeInformatica)

Example: Form Factor



Example: Use-case completion times



5. Written reports

- Varying degrees of formality, depending on purpose, audience, organizational culture. A full length formal report may include:
  - Summary
  - Introduction
  - Methods
  - Results
  - Discussion
  - Conclusions
  - Recommendations

Jordan An introduction to Usability

5. Different stakeholders, different needs

- Examples of different presentations and responses.
- Method 3: Curve Diagram. Liked by designers, can give a lot of information if interpreted properly. Ranked poorly by product owners, not interested in individual use cases.
- Method 8: PowerPoint, no verbal presentation. Liked by product owners; gives an overall view of the product in comparison to other devices, without including too much information about the context and test situation. Not useful to designers, who lack the contextual information and chance to ask follow-up questions
- Method 1: Structured Data Summary. Liked by designers because of the extent and character of the contextual information it includes, and because of the way the data is visualised. Disliked by product owners, the information is not adapted to the broad view of the product that the Product Owners need.

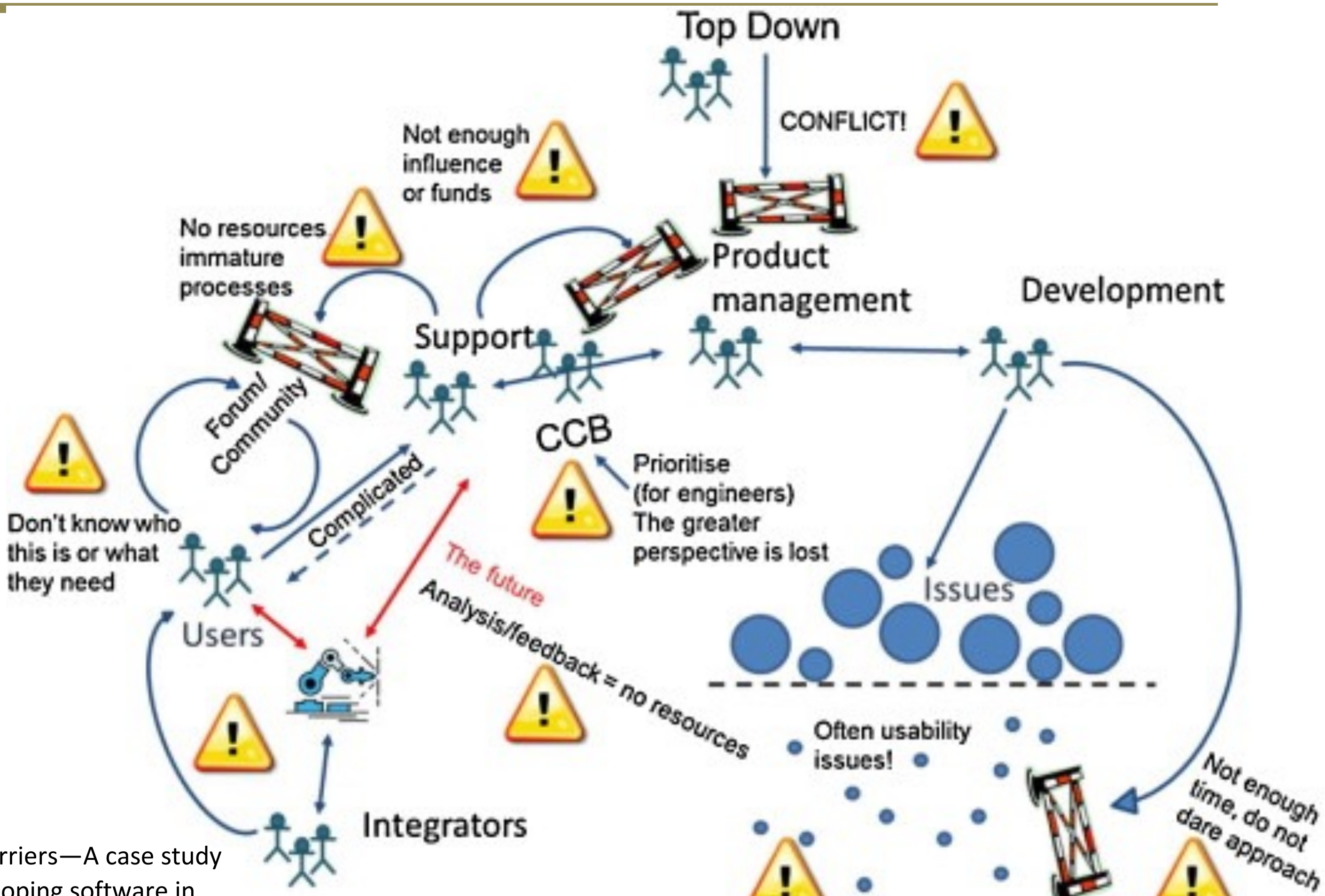
(Winter, Rönkkö, 2009)

6. The ethics of testing

- Informed consent!
- Be well-prepared
- Respect the users' emotions and well-being
  - There is pressure to perform, even if we say that the system is being tested, not the user
  - Users will make errors, or be slow to learn the system, and may feel inadequate or stupid
- Make the participants feel as comfortable as possible
- Clarify that no information regarding individuals will be released – keep all data confidential and write up results so that anonymity is maintained

Nielsen, Usability engineering, p. 181

6. Conflicting prioritizations


Top Down
CONFLICT!
Not enough influence or funds
No resources immature processes
Product management
Development
Support
Forum/ Community
CCB
Prioritise (for engineers) The greater perspective is lost
Don't know who this is or what they need
Complicated
The future
Analysis/feedback = no resources
Issues
Users
Often usability issues!
Not enough time, do not dare approach
Integrators

Identifying organizational barriers—A case study of usability work when developing software in the automation industry (2014)
J Winter, K Rönkkö, M Rissanen
Journal of Systems and Software 88, 54-73

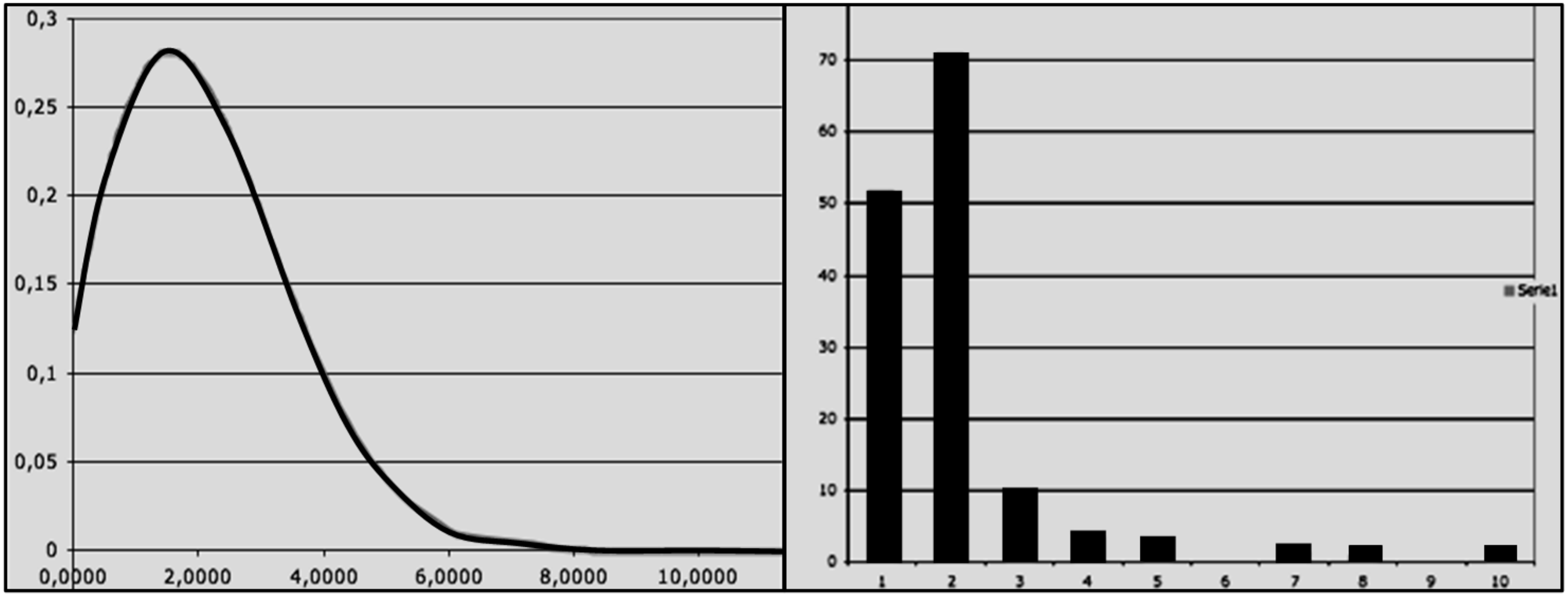
7. Further research questions

- Data can be used in many ways.
- Based on our data, we published a study regarding other ways that the data can be used:
  - RQ1: What is the correlation between the different aspects of usability (Effectiveness, Efficiency and Satisfaction)?
  - RQ2: Can a statistical analysis of task-completion time allow us to discover problematic use cases?

(Winter, Hinely, 2011)

7. Further analysis of results

- What can we learn by distribution of times taken to perform use cases?



(Winter, Hinely, 2011)

We found…

- the correlations between the factors of usability are not sufficiently strong to allow us to base usability evaluations on the basis of one single metric. This means that it is important that all three factors are measured and analyzed, and as discussed previously, the test leader is an important figure in this process.

(Winter, Hinely, 2011)

We found also…

- However, we do find that it may be possible to discover potentially problematic use cases by analysing the distribution of use case completion times. This would mean that it is possible to collect data which indicate which use cases are most important to concentrate testing resources on.

(Winter, Hinely, 2011)

Literature

- Brooke, J., *System Usability Scale (SUS): A Quick-and-Dirty Method of System Evaluation User Information*. 1986, Digital Equipment Co Ltd, Reading, UK
- Dumas, J.S. and Redish, J.C., 1999. A practical Guide to Usability Testing, Intellect
- Jordan, P.W., 2002. *An introduction to usability*. Taylor & Francis.
- Rönkkö, K., Winter, J., Hellman, M. 2009, Inside Information II, technical report, Blekinge Institute of Technology
- Winter, J., Rönkkö, K. 2009, Satisfying Stakeholders' Needs – Balancing Agile and Formal Usability Test Results, e-Informatica Software Engineering Journal, Vol 3, Issue 1
- Winter, J., Hinely, M. Examining Correlations in Usability Data to Effectivize Usability Testing, e-Informatica Software Engineering Journal, Vol 5, Issue 1
- Winter, J., Rönkkö, K., Rissanen, M., 2014, Identifying organizational barriers—A case study of usability work when developing software in the automation industry, *Journal of Systems and Software* 88, 54-73