UXQB
International Usability and UX Qualification Board

CPUX-UT Curriculum

Certified Professional for Usability and User Experience – Usability Testing and Evaluation

Version 1.07 EN, 9 February 2017

Publisher: UXQB e. V.
Contact: info@uxqb.org

www.uxqb.org

Editor: R. Molich. Co-Editor: B. Rummel. Contributors: T. Geis, O. Kluge, K. Polkehn

Content

Introduction

This curriculum defines what a student needs to know in order to pass the certification test for Certified Professional for Usability and User Experience – Usability Testing (CPUX-UT). The certification test only tests concepts, procedures and knowledge described in this document. Students must have previous knowledge about usability corresponding to the CPUX-F curriculum (available from www.uxqb.org).

The curriculum describes the minimal level of knowledge that a practitioner should have before conducting a professional usability evaluation, in particular a usability test.

The following basic requirements apply for this curriculum:

- It must conform to established conventions for usability evaluation as described in popular textbooks and international standards.
- It must conform to the CPUX-F curriculum (available from www.uxqb.org).
- A qualified teacher must be able to present the contents prudently in 3 days including appropriate exercises that take up about half of the time. See the description of the model seminar in Appendix 1.

This curriculum is not a tutorial on how to conduct usability evaluation. Readers are assumed to have some practical experience from observing or conducting professional usability evaluation. A CPUX-F certificate without practical experience may not be sufficient.

The curriculum consists of articles. Each section contains a number of related articles. Articles in a section are ordered from general to specific. Each article explains a term with a definition, notes and examples. Terms shown in boldface in an article are elaborated in a separate article. Use the index at the end of the curriculum to quickly locate an article. The definitions of some terms appear in the CPUX-F curriculum; this is shown in the index.

Acknowledgments

The Editors express their gratitude to working group members and reviewers of this and previous versions of the Curriculum for their constructive contributions to the technical contents. Particular thanks for insightful questions and valuable feedback to participants in CPUX-UT training courses.

Editor: Rolf Molich

Co-Editor: Bernard Rummel

Contributors and reviewers: Vera Brannen, Anja Endmann, Thomas Geis, Kasper Hornbæk, Rüdiger Heimgärtner, Daniela Keßner, Oliver Kluge, Knut Polkehn, Catharina Riedemann, Sabine Rougk, David Travis, Chauncey Wilson.

Legends

Grayish tables show definitions of important terms and concepts, and corresponding notes

Term	Definition

Bluish tables contain tips and recommendations

Tips	

Reddish tables tell what you should avoid

Avoid	

Learning Objectives

Learning objectives (LO) are brief statements that describe what students will be expected to learn from the curriculum.

Yellowish tables show learning objectives.

LO #	Learning Objective

The word "Foundation" in the "LO #" (Learning Objective number) column indicates that the term is defined in the CPUX-F foundation level curriculum (available from www.uxqb.org)

Learning Objectives are characterized by the keywords

Knowing – that is, reciting, recognizing

Understanding – that is, comparing, distinguishing, explaining, substantiating, summarizing

Mastering – that is, analyzing, communicating, documenting, executing, planning

1. Overview of usability evaluation

LO #	Learning Objective
Foundation	Understanding how and when usability evaluation is used in human-centred design activities as defined by ISO 9241-210 (K2)
1.1	Understanding the differences between usability test, inspection and user survey (K2)
1.2	Understanding the selection of the most appropriate usability evaluation method in a given context (K2)
Foundation	Understanding the difference between formative and summative evaluation and being able to select the appropriate approach in a given context (K2)
Foundation	Understanding the various levels of prototype fidelity and their impact on testing (K2)
Foundation	Understanding user requirements, in particular the difference between qualitative and quantitative user requirements (K2)
Foundation	Understanding the application of user requirements in usability evaluation (K2)
1.3	Understanding quality criteria for a usability evaluation (K2)
1.4	Knowing agile usability evaluation (K1)
1.5	Knowing usability maturity (K1)

Term	Definition
Usability evaluation	A process through which information about the **usability** of an **interactive system** is gathered in order to improve the **interactive system** (known as **formative usability evaluation**) or to assess the merit or worth of the **interactive system** (known as **summative usability evaluation**). Notes: 1. The purpose of a usability evaluation is to identify as many valid **usability problems** and **positive findings** as possible in an **interactive system**. 2. This curriculum discusses usability evaluation of **interactive systems**. Usability evaluation can also be applied to other products where usability matters, for example user guides, vending machines, aircraft cockpits, and the design of train stations. 3. Usability evaluation can be carried out a. with **users**; appropriate methods are **usability test** and **user survey**, b. without **users**; appropriate methods are **inspection**. Usability evaluations with **users** focus on **users**' objective abilities to solve tasks; usability evaluations without users are basically opinions based on expertise. 4. The synonym "Evaluation" is also often used. 5. Pay attention to the **quality of a usability evaluation** and **criteria for selecting a usability evaluation method**. 6. For organizations that use an agile development process, use **agile usability evaluation**.

Criteria for selecting a usability evaluation method	The criteria for selecting a **usability evaluation** method are: 1. The purpose of the **usability evaluation**. a. To increase usability awareness or **usability maturity** in the organization, run a **usability test**. This may convince skeptical **stakeholders**. b. To evaluate **effectiveness**, use a **qualitative** or a **quantitative usability test** or an **inspection**. c. To evaluate whether **efficiency requirements** have been fulfilled, or whether the **usability** of a product has measurably improved, use a **quantitative usability test**. d. To evaluate whether **satisfaction requirements** have been fulfilled, use a **user survey**. 2. **Usability maturity**. Use **usability test** rather than **inspection** if the **usability maturity** of the organization is low. Controversial outcomes from an **inspection** can be dismissed as "opinions" in an immature organization. There is no good answer to the question "Why are your opinions better than mine?" This question is an indication of lack of **usability maturity**. 3. Completion time. **Inspections** are fastest with respect to calendar time; they can often be completed within a few days. **Unmoderated usability tests** can also be completed within a few days, but they are only available for certain types of **interactive systems**, such as working websites. 4. Project stage. If the project is in the early design stage where only **low-fidelity prototypes** are available for **evaluation**, use a **formative** method such as **inspection, qualitative usability test** or **user survey**. 5. Resources. If few resources are available, consider a **discount usability test, unmoderated usability test** or **RITE**. **Usability test** requires an appropriately trained **usability tester** and **test participants**, who can be time consuming or expensive to **recruit**. **Inspections** require usability **experts** and subject matter **experts**, which can be hard to get. 6. Type of development process: Waterfall, agile or other. See **agile usability evaluation.** 7. Mix methods. It is perfectly OK to mix several recognized **usability evaluation** methods in one project. Examples: a. A quick **discount usability test** of a new app combined with a few more formal in-depth **usability test sessions**. b. A quick **inspection** that uncovers the most serious **usability problems** so they can be eliminated before an expensive **usability test**.

User experience evaluation	A process through which information about the **user experience** of an **interactive system** is gathered in order to improve the **interactive system**. Notes: 1. **Usability** is part of the **user experience**. The evaluation of the **usability** part of **user experience** is addressed in the rest of this curriculum. The techniques used for user experience evaluation are basically the same as those used for usability evaluation. 2. A user experience evaluation addresses the whole **user experience** with the **interactive system**, including for example a. Billboards and ads that make **users** aware of the **interactive system**. b. Training in the use of the **interactive system**. c. Touchpoints with the **interactive system** other than screen dialogue, for example encounters with support, letters or goods received as a result of the interaction with the **interactive system**. d. Problems that are not handled by the **user interface** of the **interactive system**, for example notifications of delays, handling of complaints, unsolicited calls, etc. 3. **User experience** can be evaluated using traditional methods. Examples: a. Bridge time gaps during a **usability test session**: "Two weeks later you receive this letter. Please open it and tell me what has happened". b. **Users** keep diaries while they use the **interactive system** over an extended time period.
Quality of a usability evaluation	The degree to which a **usability evaluation** fulfils commonly accepted requirements for professional **usability evaluation**. Notes: 1. Commonly accepted requirements for professional **usability evaluation** are a. Described in this CPUX-UT curriculum b. Described in recognized textbooks 2. An important step to improving the **quality** of **usability evaluations** is to realize that **quality** could be a problem. 3. Check for **quality** problems by a. Applying the CPUX-UT Usability Test Checklist. See “Checklist for the evaluation of the practical test” on www.uxqb.org/en/documents b. Having your approach to **usability evaluation** reviewed by one or more qualified, neutral outsiders. Reviews should occur regularly, for example once every 3 years. Reviews by colleagues are less valuable. 4. Humility and openness to constructive criticism is useful for increasing **quality**. Consider reviews and quality assurance opportunities, not nuisances. 5. **Usability testers** should pay attention to comments and criticism from peers – that is, other usability professionals, **test participants** and **stakeholders**. They might have very valid points. It's a common mistake to believe that if your **stakeholders** criticize you, it's because they don't like **usability**.

Agile usability evaluation	A **usability evaluation** approach that is suited to a development process where working but incomplete software is delivered early and frequently. Notes: 1. In an agile environment, design teams work in short development cycles, called sprints, of one week to one month. In each sprint, the goal is to get a feature or a group of features designed and coded. 2. **Usability evaluation** approaches that work well with agile development: a. **RITE**, Rapid Iterative Testing and Evaluation b. Weekly testing. **Test participants** are **recruited** well in advance and scheduled each week, for example each Tuesday, so whatever is ready can be **usability tested**. Appropriate **usability test tasks** are prepared shortly before the **usability test**. c. **Expert usability reviews** d. Quick **discount usability test sessions**, for example in a cafe, conference or trade show where you have access to lots of potential **users**. Such **discount usability test sessions** last 3-10 minutes.

Usability maturity	The level of understanding and implementing, and positive attitudes towards a systematic **human-centred design** process within an organization Notes: 1. The maturity (capability) levels of standard process assessment models according to ISO 33020 are: 0. Incomplete The **human-centred design** process is not implemented, or fails to achieve its process purpose. Note: At this level there is little evidence of any systematic achievement of the process purpose. Product managers say that they care about usability, but when it comes to spending resources or making inconvenient decisions to achieve usability, nothing happens. Usability is fine if it comes for free, but no one is committed to it. 1. Performed The **human-centred design** process achieves its process purpose. Note: Usability is achieved by enthusiastic individuals using ad-hoc processes. 2. Managed The **human-centred design** process is implemented in a managed fashion (planned, monitored and adjusted), and its work products are appropriately established, controlled and maintained. Note: The process is planned and monitored 3. Established The **human-centred design** process is implemented using a defined process that is capable of achieving its process outcomes. Note: A standard process, including appropriate tailoring guidelines, is defined and maintained. For example, the organization has a standard for how to conduct a usability test and procedures for enforcing it. 4. Predictable The **human-centred design** process is executed consistently within defined limits Note: The process is aligned with quantitative business goals 5. Innovating The **human-centred design** process is continuously improved to respond to change aligned with organizational goals. Note: Process innovation objectives are defined that support the relevant business goals 2. Level 5 is best (highest), level 0 is worst (lowest). 3. The following extension of the ISO levels is useful in practice: -1. Indifference or outright hostility towards **Usability** Developers don't care about **users** or their needs; their only **goal** is to build features and make them work on the computer. In this mindset, humans are irrelevant – they're told to use the system, regardless of whether doing so is easy or pleasant. Usability activities are considered counterproductive, suppressed and detested. 4. Usability maturity assessments of an organization are best conducted by neutral, external specialists.

2. Inspection

LO #	Learning Objective
2.0.1	Understanding when to use inspection and when to avoid it, in particular with respect to usability maturity (K2)
2.0.2	Mastering the steps of an inspection (K3)
2.0.3	Understanding heuristic evaluation (K2)
2.0.4	Understanding basic heuristics (K2)
2.0.5	Knowing quality criteria for sets of heuristics (K1)
2.0.6	Understanding usability review (K2)
2.0.7	Knowing expert usability review (K1)
2.0.8	Knowing cognitive walkthrough (K1)
2.0.9	Understanding the differences between heuristic evaluation, usability review, expert usability review and cognitive walkthrough (K2)

This curriculum breaks down an **inspection** of an **interactive system** into the following steps:

1. Prepare **inspection**
   a. Determine **goals** in cooperation with **stakeholders**
   b. Select appropriate method: **heuristic evaluation, usability review, expert usability review**, or **cognitive walkthrough**
   c. Select **evaluators**, for example usability **experts** and subject matter **experts**
2. The **interactive system** to be evaluated is presented to the **evaluators**.
3. The **evaluators** evaluate the **interactive system** in solitude and write down their **findings**.
4. The **evaluators** meet and try to reach consensus on highlights and lowlights.
   a. This consensus building is of particular importance. It weeds out any **findings** that are peculiar to a specific **evaluator**.
   b. Only **findings** that a majority of the **evaluators** support are included in the **inspection report**
   c. Step 4 is not possible if the **inspection** is carried out by a single **evaluator**.
5. The **evaluators** present the **findings** on which a majority agree to the **author** and interested **stakeholders**.
6. One of the **evaluators** writes the **inspection report**.

Term	Definition
Inspection	**Usability evaluation** based on the judgment of one or more **evaluators** who examine or use an **interactive system** to identify potential **usability problems** and deviations from established criteria. Notes: 1. A step-by-step overview of an inspection is available at the start of this section 2. Inspection-based **usability evaluation** is often performed by usability **experts** or subject matter **experts** who base their judgement on prior experience of **usability problems** encountered by **users** and their own knowledge of ergonomic **user interface guidelines** and **styleguides**. Inspections can be carried out by people with little usability knowledge, for example users. 3. Inspection techniques are: a. **Heuristic evaluation** b. **Usability review** c. **Expert usability review** d. **Cognitive walkthrough** 4. Inspection is opinion based and may cause opinion wars. An opinion war is a serious disagreement where the key arguments are unsubstantiated opinions rather than data or facts. Opinion wars are a sign of an organization that lacks **usability maturity**, or of inexperienced **usability professionals**. 5. Frequent mistakes in inspections: a. No frank discussion between **evaluators** of the **findings**. Any **finding** suggested by just one **evaluator** is accepted uncritically. It is crucial for the success of an inspection that **evaluators** are capable of rejecting questionable **findings** and recognizing good **findings** even if they are suggested by one **evaluator** only. b. **Evaluators** are not sufficiently familiar with the **interactive system** and its restraints, which means that **findings** are either not useful for development teams, or they are easily brushed aside. c. **Evaluators** focus on minor details like **user interface guideline** violations when there are serious problems with **effectiveness** and **efficiency**.
Inspection criteria	Generally accepted or agreed rules or rules of thumb for usable **user interfaces**. Notes: 1. Inspection criteria include a. **Heuristics** b. **User interface guidelines** c. **User requirements** 2. **Inspections** can be carried out without explicit criteria, for example in usability **reviews** or **expert** usability **reviews**. Such **inspections** are based solely on the **evaluators'** experience

Term	Definition
Heuristic evaluation	A **usability evaluation** method in which one or more **evaluators**, preferably **experts**, compare an **interactive system** to a list of **heuristics** and identify where the **interactive system** does not follow those **heuristics**. Notes: 1. The list of **heuristics** must be manageable. Usually about 10 **heuristics** are used. 2. **Experts** can be usability **experts** or subject matter **experts** ("single **experts**"), or both ("double **experts**"). 3. Arguments against heuristic evaluation: a. It requires the **evaluators** to make judgments by comparing a product to a limited set of **heuristics**. **Usability** is much too complex to be expressed in just 10 or even 50 **heuristics**. b. It ignores **context of use**; this makes judgment of the **interactive system** difficult, if not impossible. 4. Frequent mistakes in using heuristic evaluation: a. The evaluation is based on gut feeling rather than heuristics. **Findings** are assigned to one or more **heuristics** after the **findings** have been found. The correct approach is to let the **heuristics** drive the heuristic evaluation. b. **Evaluators** do not fully understand the **heuristics**. Many **heuristics** are compact and interpreting them correctly requires some experience. c. Experience shows that many **usability professionals** who claim that they do heuristic evaluations actually do **(expert) usability reviews**, because they report **findings** that could not possibly be found using the **heuristics**. d. The evaluators use heuristics made up by themselves instead of generally recognized heuristics.

Term	Definition
Heuristic	A generally recognized rule of thumb that helps to achieve **usability**. Notes: 1. The purpose of a heuristic is to provide reliable and useful guidance to an **evaluator** during the **usability evaluation** of an **interactive system**. 2. Criteria for reliable sets of heuristics: a. Generally recognized – a reliable set of heuristics has stood the test of time b. Comprehensible c. Useful d. Manageable – that is, not too many heuristics. Usually about 10 **heuristics** are used. Examples of heuristics: 1. According to current textbooks the most widely recognized set of heuristics are the heuristics created by Jakob Nielsen: a. Visibility of system status The system should always keep users informed about what is going on, through appropriate feedback within reasonable time. b. 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, making information appear in a natural and logical order. c. 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. d. Consistency and standards Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions. e. Error prevention Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action. f. 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. Instructions for use of the system should be visible or easily retrievable whenever appropriate. g. 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. Allow users to tailor frequent actions. h. Aesthetic and minimalist design Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility. i. Help users recover from errors Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution. j. Help and documentation Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Term	Definition
Usability review	**Usability evaluation** based on the judgment of one or more **evaluators** who examine or use an **interactive system** to identify potential **usability problems**. Notes: 1. A valuable usability review can be carried out by a **usability professional** who has 1-2 years of experience, but who does not yet qualify as a usability **expert**. 2. While some usability experience or domain knowledge is helpful, no formal usability qualifications are required for an **evaluator** in a usability review. 3. Usability reviews are often based on opinion, personal experience and common sense.
Expert usability review	A **usability review** in which the **evaluators** are **experts**. Notes: 1. Expert usability reviews are often based on extensive experience, mainly from **usability tests**, and introspection – that is, **experts** observe themselves as they carry out **tasks**. 2. **Experts** can be usability **experts** or subject matter **experts** ("single **experts**"), or both ("double **experts**"). Subject matter **experts** can be **users**. 3. It's perfectly OK to combine expert usability review and **heuristic evaluation**, for example by starting with an expert usability review and then go back over the **heuristics** later to reduce the risk that something was missed.
Expert	A person with extensive knowledge or ability based on research, experience, or occupation and in a particular area of study. Notes: 1. Gladwell's 10,000-Hour Rule claims that the key to success in any field is, to a large extent, a matter of practicing a specific task for a total of around 10,000 hours. 2. Experts are recognized as such by independent peers.
Cognitive walkthrough	**Usability evaluation** of a **user interface** in the context of one or more specific user **tasks**. Notes: 1. In a cognitive walkthrough, **evaluators** walk through the sequence of actions required by the **interactive system** for one or more **user tasks.** For each step in the user task, the evaluators consider: a. Will the **user** try to achieve the right effect? b. Will the **user** notice that the correct action is available? c. Will the **user** associate the correct action with the effect that the **user** is trying to achieve? d. If the correct action is performed, will the **user** see that progress is being made toward solution of the **task**? 2. A cognitive walkthrough is often carried out in the context of a **scenario** and a **persona**. **Evaluators** walk through a **scenario** identifying themselves with a **persona**. This kind of evaluation is sometimes referred to as a **persona** based **usability review**.
Inspection report	A document that describes the **findings** from an **inspection**. Notes: 1. The structure of an inspection report is similar to the structure of a **usability test report** 2. Inspection **findings** most often are not justified in the inspection report. Of course, if the client asks, the **evaluators** must provide a reasonable justification.

Tips	Inspection
Avoid personal comments	Discuss the product - not the people responsible.
Causes, not symptoms	Focus on causes, not symptoms. Example: Symptom: Users don't know how to proceed after they have completed a form Cause: A styleguide is missing or it is not being observed.
Error handling	Check error handling carefully. Try to break it.
Inspection vs. usability test	The choice depends on: 1. Availability of real **experts** (single, double) 2. Availability of **users** 3. Confidentiality of **interactive system** to be **evaluated** 4. **Usability maturity** in the organization. See also **Criteria for selecting a usability evaluation method**. 5. Available resources. In general, **inspections** are more cost-effective than **usability tests**.
Positive findings	Positive findings are just as important in the results of an inspection as they are in the results of a usability test. Please refer to section 3.3, "Positive finding".

2.1. Roles in an inspection

LO #	Learning Objective
2.1.1	Understanding the evaluator role in an inspection (K2)
2.1.2	Knowing the author role (K1)

Term	Definition
Evaluator	A person who carries out an **inspection**. The evaluator must not be an author.
Author	A person who develops or maintains the **interactive system** that is the subject of an **inspection**.

3. Usability test

LO #	Learning Objective
3.0.1	Mastering the 3 phases and the 4+5+4 steps of a usability test (K3)
3.0.2	Understanding the importance of involving stakeholders in the planning and execution of the usability test (K2)
3.0.3	Understanding the difference in purpose between a qualitative and a quantitative usability test (K2)
3.0.4	Understanding how to usability test products like mobile phones, ticket vending machines, TV sets, apps, and collaborative applications (K2)

This curriculum breaks down a **usability test** into the following 3 phases and 4+5+4 steps. Each step is described in more detail in this section.

1. Prepare **usability test**
   a. **Usability test plan**
   b. **Usability test script**
   c. **Pilot usability test session**
   d. **Recruitment** of **test participants**
2. Conduct **usability test sessions**
   a. Preparation of session
   b. **Briefing** – Pre-session instructions
   c. **Pre-session interview**
   d. **Moderation**
      - **Usability test tasks**
   e. **Post-session interview** – Debriefing
3. **Communicate findings**
   a. **Analysis of findings**
   b. **Selling findings**
      - The **KJ-method** – **Affinity diagramming**
   c. **Usability test report**
   d. **Video summary**

Term	Definition
Usability test	A **usability evaluation** that involves representative **users** performing specific **tasks** with the **interactive system** to enable identification and analysis of **usability problems,** or the measurement of **effectiveness**, **efficiency**, and user **satisfaction**. Notes: 1. A step-by-step overview of a usability test is available at the start of this section 2. Usability tests are conducted for the following major reasons: a. To assess whether user requirements have been met b. To uncover **usability problems** so they can be corrected c. To demonstrate convincingly to **stakeholders** that their **interactive system** contains serious **usability problems**. For more information see section 3.3, in particular the article **Selling findings**. 3. Results from a usability test are unique in one aspect: They show what representative **users** are able to accomplish with the **interactive system** when they carry out representative **tasks**. Eliciting personal opinions from **users**, or discussing them, does not support this objective and should be left to other methods. Examples: a. **Test tasks** like "Is the design of the home page appropriate for the CD-Shop?" are opinion based and thus inappropriate for a **usability test**. b. Remarks like "I can do this easily but most others will have serious problems" from a **test participant** are personal opinions. The **moderator** may obtain additional, valuable insight by following up on this remark with the question: "Why do you think so?" c. It's OK to report **findings** that are based on opinions about an **interactive system**, for example "The design of the home page is really pretty", if they are expressed spontaneously by a substantial number of **test participants**. 4. The concept "usability test" usually refers to a test where the **test participant** and the **moderator** are in the same physical location (“face-to-face usability test”). Other forms of usability tests are **remote usability test** and **unmoderated usability test**. 5. A usability test can be a **qualitative usability test** or a **quantitative usability test**. 6. A usability test may take place at any time during **human-centred design**, from early analysis through **interactive system** delivery and beyond. A usability test may be based on a **prototype**, in the form of paper sketches or display mock-ups, as well as on **interactive systems** under design and completed **interactive systems**. 7. A usability test is conducted by one or more **usability testers**. For other **roles** in a usability test, see section 3.4 8. A qualitative usability test is sometimes referred to as **think aloud** test. 9 In a usability test, the **moderator** works with human beings. The **moderator** must observe the **ethical rules** for working with **test participants**.
Qualitative usability test	A **usability evaluation** that involves representative **users** performing specific **tasks** with the **interactive system** to enable identification and analysis of **usability problems,** focussing on understanding **user needs**, **goals** and reasons for the observed **user** behavior. Note: 1. If the primary purpose of a **usability evaluation** is to obtain figures for the **effectiveness** or **efficiency** of the **interactive system** then the test is not a qualitative but a **quantitative usability test**.

Tips	Usability tests in general
Involve stakeholders	Involve **stakeholders** in the planning and execution of a **usability test**. This serves to make them feel that the **usability test** is also "their baby" and makes acceptance of **usability problems** smoother. The most important **stakeholders** are the people who decide which changes from the **usability test** should actually be implemented, and the people who do the actual implementation. Some ways of involving **stakeholders** are: 1. Invite them to participate in writing the **usability test plan** 2. Invite them to participate in writing the **usability test script**, in particular the **usability test tasks** 3. Invite them to participate in the **recruitment** process, in particular defining the **test participant** profile and creating the **recruitment screener** 4. Invite and encourage them to observe **usability tests**. 5. Make it easy for them to observe **usability test sessions**. a. Schedule **usability test sessions** at times that are convenient for **stakeholders**, for example on Friday afternoons. Advertise **usability test sessions** widely, and indicate that it's OK to observe only part of a **usability test session**. b. Carry out **usability test sessions** at locations that are convenient for **stakeholders**, for example at a place where many **stakeholders** work instead of at a far away **usability lab**. c. Make it easy for **stakeholders** to observe **usability test sessions** as a group, because seeing is believing and watching and discussing makes converts. Many good effects flow from watching as a group. 6. Use the **KJ-Method** to actively involve **stakeholders** in the **analysis of findings**.
Resource estimate for usability test	1. A rough resource estimate for a complete **moderated, qualitative usability test** with 5 **test participants** and 45-minute **usability test sessions** is 50 to 80 person hours. 2. A complete **moderated, qualitative usability test** with 5 **test participants** and 45-minute **usability test sessions** can be done in about 2 weeks from start of writing the **usability test plan** to end of **communication of findings**. 3. Reliable estimates for **quantitative usability tests** are currently not available.

3.1. Prepare usability test

LO #	Learning Objective
3.1.1	Mastering the usability test plan (K3)
3.1.2	Understanding the factors that influence the number of test participants used in a usability test (K2)
3.1.3	Mastering the usability test script (K3)
3.1.4	Mastering recruitment including the recruitment screener and the confirmation process (K3)
3.1.5	Understanding the non-disclosure agreement and the informed consent agreement (K2)
3.1.6	Understanding the factors that influence the choice of the incentive (K2)
3.1.7	Understanding the purpose of a pilot usability test session (K2)

3.1.1. Usability test plan

Term	Definition
Usability test plan	A short description of the purpose and extent of a **usability test**. Notes: 1. The usability test plan is intended for management or a client to decide whether the **usability test** should be run or not. 2. Start preparations for a **usability test** by writing a usability test plan. Ask **stakeholders** and management to review the plan and modify it until it's acceptable. 3. The usability test plan includes a. The **goals** of the **usability test;** b. The **user group** or user groups for the **usability test**; c. A reference to the **user requirements** for the **interactive system** (if the purpose of the **usability test** is to evaluate the design against **user requirements**); d. Number of planned **test participants**; e. Approximate length of each **usability test session**; f. Name of **moderator** or moderators; g. Time plan; h. A **resource estimate** for the **usability test**; i. How the **findings** will be communicated 4. The usability test plan must be short and to the point. Often, one page suffices. 5. Further details about the **usability test** such as **usability test tasks**, test method and required software and hardware are provided in the **usability test script**.

3.1.2. Usability test script

Term	Definition
Usability test script	A checklist used by a **moderator** in a **usability test** to keep track of a. The **preparation** for the **usability test session** – that is, what to do before the **test participant** arrives, b. **Briefing** instructions, c. **Pre-session interview** questions, d. **Usability test tasks**, e. **Post-session interview** questions. Note: 1. A sample **usability test script** is included in the **sample usability test report**.

3.1.3. Pilot usability test session

Term	Definition
Pilot usability test session	A **usability test session** that is conducted in accordance with the **usability test script** in order to test the **usability test script** and the usability test setup. Notes: 1. If serious problems in the **usability test script** are discovered in a pilot usability test session, correct the **usability test script** and conduct another pilot usability test session. 2. Pilot usability test sessions are particularly important for **unmoderated usability tests** because no **moderator** is present to intervene if the **test participant** misinterprets a **usability test task**. 3. Prefer real **users** as **test participants**. If they're not available, use colleagues. Avoid using people who designed the **interactive system** as **test participants**. 4. **Findings** from pilot usability test sessions may be included in the **usability test report** if the **findings** seem valid. For example, if the **test participant** in the pilot test session is over-qualified for the **usability test** and still encounters a serious **usability problem** in a valid **usability test task**, this **finding** may be considered valid.

3.1.4. Recruitment of test participants

Term	Definition
Recruitment	A process for selecting candidates that have the required qualifications to participate in a **usability test**. Notes: 1. A **recruitment screener** is often used to determine whether candidates have the required qualifications to participate in the usability test. 2. **Test participants** from previous **usability tests** can be re-used as long as you are aware of the bias this may introduce. Consider re-using **test participants** who are good at **thinking aloud** and provide frank and apt comments, in particular for **usability test sessions** that are observed by important **stakeholders**. Avoid people who have participated in so many **usability tests** that they are no longer representative of their **user group**. 3. When you recruit in-house colleagues who participate in a **usability test** during normal work hours as part of their job make sure to inform or get permission from the manager of the prospective **test participant** ahead of the recruitment. 4. Consider using a professional recruiting agency. Check the quality of the agency by asking them to call you in an attempt to recruit you. 5. Experience shows that 10 to 20% of **test participants** will be "no-shows" – that is, they will not show up. Some may not even bother to cancel the appointment. This happens more frequently if the **incentive** is low and you have no relationship to the **test participants**. If you recruit in-house colleagues, no-shows are rare. For critical **usability test sessions**, consider recruitment of a "floater" – that is, a spare **test participant** who is used only if the regular **test participant** does not show up. Floaters get paid for doing whatever they want if all **test participants** show up. Tip: Some floaters get disappointed if they are not allowed to be **test participants** because all regular **test participants** show up. If this happens, run an additional **usability test session** with the floater for ethical reasons. 6. Temp agencies can often recruit **test participants** from their lists of temporary workers. Temporary workers are often able to participate in **usability test sessions** that take place in the middle of the day when most other people are at work. Also, temporary workers most often show up exactly as agreed. The downside to this is that some temp agencies will charge you for several hours even if the **usability test session** only lasts 1 hour. If you use a temp agency for recruitment, make sure that **test participants** have the required qualifications. Be careful not to simply recruit people who are available. 7. Market research companies are great if you want **test participants** with special qualifications. 8. Explicitly inform **test participants** during recruitment if you intend to video or audio record the **usability test session**.
Recruitment screener	A series of questions for prospective **test participants** to identify whether they represent the intended **users** and therefore qualify to participate in a human-centred activity, for example a **usability test** or a **focus group**. Notes: 1. Relevant qualifications include: Background, knowledge of the subject matter, attitudes towards the subject matter, expertise with technology, and interests. Age and gender are often of less importance, even though they are sometimes related to relevant qualifications; for example age is sometimes related to attitudes towards technology. 2. Exclusion criteria for **test participants** are: User experience or design professionals, journalists, people employed by competitors. 3. Ask all questions that may cause a prospective **test participant** to be rejected as early in the screener as possible. 4. Have a diplomatic phrase ready when you must reject prospective **test participants** that do not qualify for a study, for example "Your qualifications are great. They're just not what we're looking for at this time" or "I'm afraid you are over-qualified for this study."

Term	Definition
Informed Consent Agreement (ICA)	A written agreement that informs the **test participant** of his or her rights and obligations in connection with a **usability test**. Notes: 1. Use a **non-disclosure agreement** if the information disclosed to the **test participant** during the **usability test session** is confidential. Otherwise use an informed consent agreement. 2. An informed consent agreement (ICA) should contain: a. Purpose of **usability evaluation** b. "The usability evaluation session will be video recorded" (if this is indeed the case) c. "The usability evaluation session will be observed from a neighboring room by people who are actively involved in developing the product" (if this is indeed the case) d. "We are not evaluating you. We are evaluating the product" e. "You are free to leave any time. Even if you leave early you will still get your gift" f. "If you need a break, just tell us.". 3. A data privacy statement may be required by law or by local regulations to inform **test participants** about what happens to data from their **usability test session**, for example video recordings and observer notes. 4. **Test participants** are expected to sign the agreement at the start of the **usability test session** before any information about the **interactive system** is disclosed to them. 5. Ensure that the agreement is usable. An agreement is worthless if **test participants** do not understand what they are signing.
Non-Disclosure Agreement (NDA)	An extended **informed consent agreement** that is used if the information disclosed to the **test participant** during the **usability test session** is confidential. Note: 1. An NDA should contain the following information in addition to the information in the **ICA**: a. "You will be working with a product that is still being developed. Any information you acquire about this product is confidential and proprietary and is being disclosed to you only so that you can participate in the evaluation. By signing this form, you agree not to talk about this product to anyone. You may tell them that you helped to evaluate new software."

Term	Definition
Incentive	A gift or payment given to a **test participant** as a "thank you" or compensation for participating in a **usability test**. Notes: 1. Make the incentive feel like a “thank you” gift or a reasonable compensation for the **test participant**’s time and not a bribe to influence what they will be saying about the **interactive system**. Traditionally, incentives are handed out after the **usability test session** has been completed. This could be wrongly interpreted as if the **test participant** is actually being evaluated – that is, the incentive is a reward for pleasing the **moderator**. The **moderator** can avoid this by handing out the incentive at the start of the **usability test session** while he or she is saying "we are not evaluating you". 2. Check if your employer or client or employer of **test participants** has price limits for gifts in order to avoid even the appearance of bribery. This is particularly important if **test participants** are publicly employed. 3. Avoid incentives that could be perceived as insulting, for example coffee mugs with a company logo. Do not indicate in any way that you consider your **test participant**'s time less valuable than your time. 4. If you hand out cash incentives, consider tax regulations. 5. If a receipt is required, have the appropriate form ready for the **test participant** to sign. 6. Typical incentives at the time of writing in Northern Europe are on the order of 50€ for a session of about one hour in cases where no special qualifications are required. For **test participants** with special qualifications, for example lawyers or doctors, it may be up to 10 times higher, and **test participants** may demand that the **usability test session** is carried out at their place. 7. You can avoid having to handle incentives yourself by doing the **recruitment** through a temp agency.
Confirrmation to test participant	Short, usable information that is sent to the **test participant** ahead of the **usability test session** by email or letter. Notes: 1. The confirmation should include: a. Date and time of the **usability test session**. b. Location of the **usability test session** including usable information about how to get to the location. c. "We are not evaluating you. We are evaluating the product." d. Name and contact information of contact person in case the test participant gets delayed or has difficulties finding the test location. e. Explain briefly what will happen during the **usability test session**, for example “You will help us evaluate a new product. I will ask you to carry out a number of tasks with the new product. I will also ask you to answer a number of questions.” 2. Avoid any use of terms that might make **test participants** uneasy about the **evaluation**, like "lab", "laboratory" or "experiment". Calm **test participants** by informing them that evaluations like this are routine. 3. Make sure that the information is usable. It should be short and precise. Listen carefully to feedback from **test participants** or **usability test** the confirmation. 4. Explicitly inform the **test participant** if you intend to video or audio record the **usability test session**. 5. Consider including the **non-disclosure agreement** or the **informed consent agreement** with the confirmation so the **test participant** does not have to spend valuable time during the **usability test session** reading and signing it.

Tips	Recruitment
Number of test participants	For a **qualitative usability test**, even one or a few **test participants** can provide important insight as exemplified by the **RITE** method. A rule of thumb is that 5 **test participants** are enough to drive a useful iterative cycle. In other words, once you have conducted 5 **usability test sessions**, stop, have the **usability problems** corrected, and run 5 additional **usability test sessions** on the corrected **interactive system** if you have the resources. This is more efficient than using all your resources to run 15 **usability test sessions** on the same version of the **interactive system**. Ideally, all 5 **test participants** should belong to the most challenging **user group** for the **interactive system**. Finding all or even 50% of all **usability problems** in an **interactive system** is difficult. Unfortunately, it is a widespread misunderstanding that 5 **test participants** will find 75% or more of the **usability problems**. For a **quantitative usability test**, the number of required **test participants** starts at 20 because of statistical uncertainty. Please see the discussion in section 3.5, in particular the article about **confidence interval**.

3.2. Conduct usability test sessions

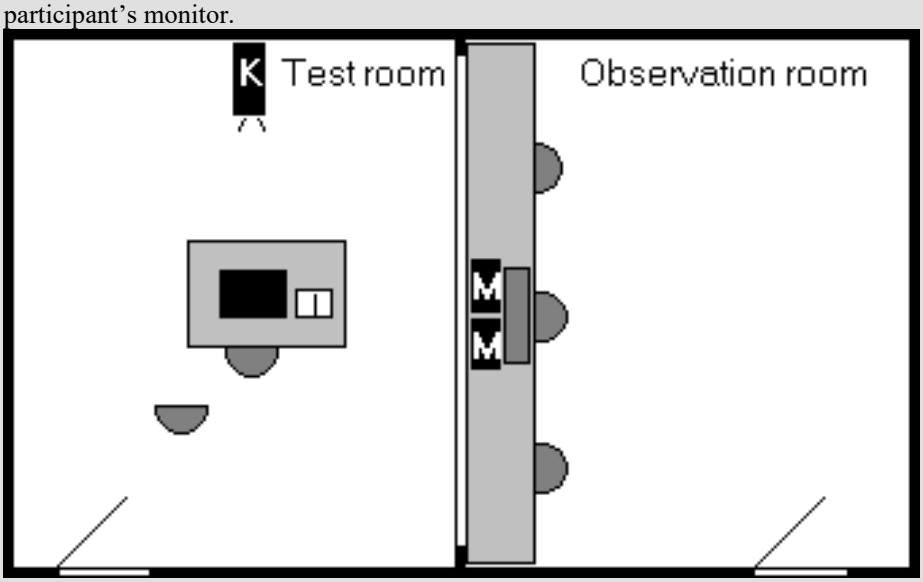
LO #	Learning Objective
3.2.1	Mastering the flow of a usability test session (K3)
3.2.2	Understanding the criteria for the selection of a suitable test location (K2)
3.2.3	Knowing the usability lab (K1)
3.2.4	Understanding the preparation of a usability test session (K2)
3.2.5	Mastering the briefing (K3)
3.2.6	Mastering the pre-session interview (K3)
3.2.7	Mastering the moderation (K3)
3.2.8	Understanding typical problems in moderation (K2)
3.2.9	Mastering usability test tasks (K3)
3.2.10	Understanding typical problems in usability test tasks (K2)
3.2.11	Mastering the post-session interview – debriefing (K3)

Term	Definition
Usability test session	A part of a **usability test** where one **test participant** carries out representative **usability test tasks** using the **interactive system** or a **prototype** of the **interactive system**. Notes: 1. A **usability test** consists of a number of **usability test sessions**. 2. Usually, **usability test sessions** are moderated by a **moderator** and observed by a number of **observers**, who are often **stakeholders**. A **note-taker** records important **findings**. The moderator is the only person who is allowed to talk to the test participant during the usability test session. For additional **roles** in a **usability test** see section 3.4. 3. In a usability test session, the **moderator** typically a. Prepares the usability test session as described in section 3.2.1 b. Greets the **test participant**; c. Conducts the **briefing** d. Conducts the **pre-session interview**, e. Hands out **usability test tasks** to the **test participant** f. Observes the **test participant** during **usability test task** solution, g. Conducts the **post-session interview**. 4. The length of a usability test session is anywhere from 5 to 90 minutes. Longer sessions may fatigue or stress the **test participant** and the **moderator**. The length of a moderated usability test session is often 45-60 minutes. The length of an **unmoderated usability test** session is often 15-20 minutes. 5. **Moderators** should not moderate for more than 8 hours per day including breaks of at least 15 minutes between usability test sessions. 6. A usability test session with two or more **test participants** is called **co-discovery**.

3.2.1. Preparation of usability test session

Term	Definition
Preparation of usability test session	The activities that happen before the **test participant** arrives for a **usability test session**. Notes: 1. Checklist for the preparation: a. Set the screen resolution and zoom level to agreed conditions b. Clean up the desktop – no unnecessary application or document icons c. Lock the task bar. The task bar should be visible so time appears on video d. Turn off unwanted applications such as mail, softphone, automatic backup service e. Set up or reset the equipment to match the starting condition defined in the **usability test script**. This includes deleting any data from the previous **usability test session**, including cookies f. Put the cards with the **usability test tasks** in the right order. Each task should be printed on a separate card g. Clear the browser cache 2. Start preparations early. A useful guideline is "It's better that you wait 5 minutes for the **test participant** than that the **test participant** has to wait 2 minutes for you to finish the preparations." 3. Consider video recording. A simple webcam or screen recording tool can be useful. Trial versions of usable software screen recorders are available for free, for example from TechSmith. More expensive, professional tools such as Morae and Noldus may allow bookmark definition, keystroke and mouse logging, timestamps in the video, etc. Before deciding to video record a usability test session, decide what you want to use the recording for. Watching recorded **usability test sessions** is time-consuming and often not cost-effective.

Term	Definition
Test location	The place where the **usability test** is conducted. Notes: 1. Rules of thumb: a. **Observers** and **test participants** should be in separate rooms. If this is not possible, observe the restrictions in note 6. b. The test location should be undisturbed unless it is part of the **usability test** that the **test participant** is deliberately disturbed. c. **Observers** and **note-takers** should be able to observe what the **test participant** does without disturbing the **test participant**, either through a one-way mirror in a **usability lab**, or through video in a separate observation room. d. The **moderator** can sit with the **test participant** (recommended) or in the observation room. If the **moderator** sits with the **test participant**, the **moderator** should sit next to the **test participant** and slightly behind the **test participant**. The **moderator** must be out of sight when the **test participant** looks at the screen but it must still be possible to have eye contact. 2. Examples of test locations: a. **Usability lab**. b. Two office rooms that are connected by a video link. c. An office room. e. A room at the place where the **test participant** lives or works. f. A public place, for example a cafe. This choice is most often used for **quick discount usability test sessions** lasting 10 minutes or less. 3. It is an important part of a **usability test** that **observers** can discuss, laugh and cry while they observe a **usability test session**. Also, **observers** should be able to come and go as they please. Regulations limiting the behavior of **observers** in an observation room are a sign that the usability test setup is unusable because it does not consider the needs of the primary users. The primary users of any **usability test session** are the **observers**. 4. It is technically possible to let **observers** observe **test sessions** from their work place using screen sharing. This option is not desirable because an important part of observing a **usability test session** is the interaction and discussion with other **stakeholders**. 5. If **stakeholders** are invited to observe **usability test sessions**, choose a test location that is as convenient for the **stakeholders** as possible rather than a remote **usability lab**. 6. **Observers** and **note-takers** should be in one room, and **test participants** should be in another room. If this is not possible, for example because the test takes place in a public place or only one room is available, **test participants** may be aware of the **observers**. In this case, limit the number of **observers** to one or two and instruct the **observers** to a. Keep completely quiet during the **usability test session** unless addressed directly by the **moderator** b. Sit in a place where they are out of sight of the **test participant**. This would usually be behind the **test participant**. c. Arrive before the **test participant** arrives d. Stay in the room until the **test participant** has left 7. It is important that **observers** can hear clearly what the **test participant** says. This applies particularly for **remote** and **unmoderated usability tests**. In an **international usability test** where simultaneous translation is used, the translator must have access to high quality undisturbed sound.

Term	Definition
Usability lab	Two or more rooms that are specially equipped for **usability test**. Note: 1. A usability lab often consists of a. A test room where the **test participant** sits in front of the display unit. Often, the **moderator** sits close to but out of view of the **test participant** as illustrated by the chair to the left behind the **test participant**'s chair in the Figure. b. An observation room where **observers** (**stakeholders**) and **note-takers** can watch **test participants** as they solve **usability test tasks**. Often, the two rooms are separated by a one-way mirror which enables **observers** to watch the **test participant** but not vice versa. The Camera in the test room (K in the Figure) is connected to a monitor (M) in the observation room. It enables **observers** to see the **test participant**'s face. This camera can be a webcam located on the test participant's monitor.  2. The advantages of using a usability lab are that a. **Observers** can observe **usability test sessions** together, b. **Usability test sessions** are easy to observe for **note-takers** and **observers**, c. **Usability test sessions** are conducted under similar conditions, d. **Usability test sessions** are easy to video record, e. **Observers** can enter and leave during **usability test sessions**, f. It's a great showpiece for management who wants to demonstrate physically to visitors that the organization is deeply committed to usability 3. The disadvantages of using a usability lab are that a. The context is artificial b. It is expensive to set up and maintain.

3.2.2. Briefing – Pre-session instructions

Term	Definition
Briefing	The first activity in a **usability test session** where the **test participant** is informed about the purpose of the **usability test** and what their role and contribution are. Notes: 1. Checklist for the briefing: a. Turn off your mobile phone – and ask everyone in the room to do so, too b. Briefly introduce yourself and any other people who are in the room c. Ask **test participants** to read and sign the **non-disclosure agreement** or the **informed consent agreement**. Tip: Send the agreement to **test participants** ahead of the **usability test session** so you don't lose valuable time waiting for them to read and sign it. d. Ask for permission to video record the **usability test session** (even if the permission is part of the **NDA** or **ICA)**. e. When you have the permission and the signed agreement, start the recording. Continue to record until the **test participant** has left the room. Even after the **post-session interview** has ended, interesting remarks may come up. f. Explain the purpose of the **usability test**. g. Explain any unusual equipment in the room, for example microphones, video cameras and one-way mirrors h. Say "We are not evaluating you. We are evaluating the product" i. Say "You are free to leave any time. Even if you leave early you will still get your gift" j. Briefly explain that we want the **test participant** to **think aloud**. Don't spend too much time on this because you can't teach a **test participant** to **think aloud** in a short demonstration, and you get useful results even if the **test participant** doesn't **think aloud**. k. Finish the briefing by asking "You are welcome to ask questions at any time. Do you have any questions right now?" 2. A briefing should take at most 5 minutes. If it takes longer, chances are that the test procedure should be simplified since it takes so long to explain it, or the **moderator** is too talkative, which means that the **usability test session** is not cost effective.

3.2.3. Pre-session interview

Term	Definition
Pre-session interview	An activity in a **usability test session** where the **test participant** answers questions about his or her background and previous experience with the **interactive system** or related **interactive systems**. Notes: 1. Pre-session interview questions should briefly address the **test participant's**: a. Personal background. Often, the profession is of some interest, while age is rarely relevant. b. General knowledge of technology. Examples: "Do you have a smartphone?", "What do you use it for?", "Please explain to me how you can remove an app from your smartphone", "What would you do to get an app on your smartphone?" c. Previous experience with the **interactive system**. Examples: "Are you familiar with this system?" If the answer is yes: "What have you used it for?", "What was your experience with it?", "Please tell or show me in some detail what you did." d. Previous experience with similar **interactive systems**. Examples "Are you familiar with similar systems?" If the answer is yes: "Which one?", "What have you used it for?", "What was your experience with it?", "Please tell me in some detail what you did." 2. The pre-session interview should be brief. Do not ask for information that you already have from a reliable recruiter. Only ask for information that will really make a difference in your analysis. 3. Ideally, the pre-session interview should be part of **recruitment**. After the **test participant** has answered the pre-session interview questions during **recruitment**, submit the answers to the client for approval of the **test participant**. 4. If there's no pre-session interview during the **usability test session**, provide a summary of the **test participants**' background to **observers** ahead of the **usability test sessions**. 5. If the pre-session interview shows that the **test participant** does not match the target profile so that a test would not make sense, abort the **test session**. If you are working with a professional recruiting agency, you should have clarified beforehand how to handle such cases.

3.2.4. Moderation

Term	Definition
Moderation	The activity carried out by a **moderator** in a **usability test**. Notes: 1. During moderation the **moderator** a. Presents **usability test tasks** one by one to the **test participant**. To avoid misunderstandings, it's best to give **test tasks** in writing. In rare cases, where the spelling of a word in the task may provide a clue to solving the task, present the task verbally. **Test task** wording should be concise and clear, to minimize effort to read and understand them. Do not disclose the overall number of **test tasks**; this might upset **test participants** who only finish a subset. In general, present test tasks in the same order to all test participants. It may be OK to vary the order, for example to ensure that test tasks that are late in the sequence are also tested. Drop a test task if after some test sessions it provides no insight, or if the associated usability problems are totally clear. b. Asks the **test participant** to describe in their own words what the **test task** is asking them to do. With this, the **moderator** can check whether or not the **test participant** fully understands the **test task**. If necessary, correct the **test participant's** understanding politely (mind that probably your formulation was unclear). c. Asks the **test participant** to start working on the **test task**. d. Observes the **test participant** during **test task** solution, e. Guides the **test participant** if he or she gets completely stuck, usually by moving on to the next **test task**. Helping the **test participant** to complete a **test task** that causes serious difficulties often provides **inadvertent clues**; such help should be avoided unless solving the current **test task** is necessary to solve the following **test task**. If the **moderator** provides help, the help must be limited to an absolute minimum. f. Guides the **test participant** if he or she strays away from the **interactive system**, for example to another website that is of no interest to the current **usability test**. 2. The **moderator** should say as little as possible during moderation. Neutral utterings that indicate that the **moderator** is paying attention like "OK" and "aha" are acceptable. 3. Avoid any kind of instructions in connection with **usability test tasks**; they may provide inadvertent clues. 4. The **moderator** should be the only person who talks to the **test participant** during moderation. Other people may talk to the **test participant** only after being explicitly asked to do so by the **moderator**. 5. The most important moderation problems are: a. **Talkative moderator**, b. **Inadvertent clues**, c. Lack of curiosity, d. **Moderator** confirms solution prematurely before the **test participant** clearly states that they feel they have succeeded. e. Moderator reaches across the test participant for the keyboard or the mouse to enter data or bring the interactive system into a desired state 6. "Facilitation" is sometimes used as a synonym for **moderation**.

Tips	Moderation
Think aloud	It is useful if **test participants** share their thoughts with others during a **usability test session** by thinking aloud. Think aloud is useful for the **moderator**, **note-taker** and the **observers** to understand the **test participant**'s thoughts, **mental model** and vocabulary during a **qualitative usability test session**. In a **quantitative usability test** that measures for example **task completion time**, think aloud should be discouraged as it may influence the measurements. During the **briefing**, encourage the **test participant** to think aloud. If the **test participant** does not think aloud, remind them once or twice. If that doesn't help, leave them alone. Usually, the actions they attempt on the **interactive system** are more revealing than what they say. Avoid "relaxed think aloud" in which the **moderator** asks **test participants** for explanations and comments while they perform **test tasks**. **Test participants** should be encouraged to think aloud, not to reflect on what they are doing.
Respect	The **moderator** must be respectful towards the **test participant, stakeholders** and the client. The **moderator** must pay attention to **test participant** suggestions even if they appear unreasonable. The **moderator** must not badmouth the **interactive system** or the development team, even if it is done to calm the **test participant**. Example: A **moderator** is running a **usability test session**. The whole product team is watching. The **test participant** is struggling and finally says "You must really think I am dumb since I am unable to solve this problem". A respectful answer is "Everything you have done until now makes perfect sense to me. It's just not what the designer intended."
Co-discovery	A **usability test session** may be conducted with two **test participants** who carry out **usability test tasks** together. This is called "co-discovery". Co-discovery makes **think aloud** unnecessary because talking to each other comes naturally for **test participants**. However, co-discovery requires twice as many **test participants**, and some people find working with strangers annoying especially if their problem solving techniques differ, so despite the advantages co-discovery is rarely used in practice.

Avoid	Moderation
Talkative moderator	If the **moderator** talks too much, the **usability test session** turns into an interview, a friendly conversation or even an interrogation. Everyone feels comfortable, but the **usability test session** is not cost effective, and the **moderator** may provide **inadvertent clues** about the **interactive system**. The main purpose of a **usability test** is to observe the **test participant**, not to provide a forum for the **moderator**. This, of course, should not prevent the **moderator** from asking short, reasonable and clarifying questions regarding the **test participant**'s actions.

Inadvertent clues	Unintended hints provided by the **moderator** that help a **test participant** to solve a **usability test task**. While the **test participant** is carrying out **usability test tasks**, the **moderator** should watch his or her spoken language or body language in order not to provide unintended hints about how the **test task** is solved or the progress of the **test participant**. Examples of clues: 1. ”Don’t worry about this” signaling "even though you may think so, this is not important for this task" 2. A faint smile or the **moderator**'s tone of voice may signal "You have almost solved the task. The answer is on the screen.” 3. The **user interface** uses a term that is unknown to the **test participant**. The **moderator** uses the term in such a way that the **test participant** understands the term from the **moderator**’s remarks. 4. A **test participant** can't solve a **usability test task**. After giving up, the **test participant** asks the **moderator** to show them how the task could have been solved. The **moderator** must refrain from doing this since the **moderator**'s solution may provide clues to how subsequent **test tasks** can be solved. If the **test participant** insists, which rarely happens, show them how the **test task** could have been done during the **post-session interview**.
Leading question	The **moderator** must not ask leading questions – that is, questions that signal a preference for certain possibilities or attempt to direct the reply in a certain direction. Note: 1. Avoid “confirmation questions” where the moderator asks the test participant to confirm an answer. Confirmation questions considerably reduce the opportunity for surprises. See example 4. Examples of leading questions: 1. "Would you have liked to get special offers? For example, a special price or free accessories?" Instead ask "What do you think of the purchasing procedure?" 2. ”Would you have preferred to have the option of only looking for rental cars with automatic transmission?” 3. Any question that starts with “Would you expect ...”, for example “Would you expect the message to appear on top?” 4. “Do you always do it like this?”
Bias	The **moderator** must be unbiased and neutral. Notes: 1. The **moderator** must not defend the **interactive system**. 2. The **moderator** must not express their own views on the **interactive system**. 3. The **moderator** must not have a secret agenda such as a pet peeve or theory about the **user interface** being tested and try to get the **test participant** to articulate it. 4. Avoid using people who designed the **interactive system** as **moderators**.

3.2.5 Usability test tasks

Term	Definition
Usability test task	A description of a **task** that a **moderator** asks a **test participant** to carry out during a **usability test**. Notes: 1. A good test task a. Matches the **goals** of the **usability test** as defined in the **usability test plan**. b. Is relevant from the **test participant**'s point of view. Avoid **system oriented usability test tasks**. c. Is relevant from the **stakeholders**' point of view. 2. For each usability test task describe a. The precise phrasing of the task or **usability test task scenario** handed to the **test participant**. b. Preconditions for the task, including what's available to the **test participant** and the exact starting point. The ending point for a task is not always a valid starting point for the following task. c. Why the task was deemed to be important for the **evaluation**, including what the task is supposed to evaluate d. Any data given to the **test participant** for solving the task, for instance a delivery address, or information in the database when the **test participant** starts the task e. Criteria for task completion or task abandonment. This includes the intended outcome or expected answer. A sample criterion for task abandonment is "If the **test participant** hasn't found an answer within 10 minutes, the task is abandoned" 3. Test **key tasks** before testing specialized tasks. 4. Usability test tasks must be consistent. 5. Usability test tasks must be given in an order that seems logical from the **test participant**'s point of view. For example, ask **test participants** to order something before you ask them to cancel an order. If possible, avoid tasks that depend on the successful completion of a previous task. 6. Prepare so many usability test tasks that you are sure that you won't run out of tasks within the time allocated for the **usability test session**. 7. The first usability test task should be simple so **test participants** experience a quick success. This is particularly important if a **test participant** appears a bit stressed by the situation. The last task should also be simple to not let the **test participant** go with a negative and frustrating experience. Unfortunately, it is not always possible to keep the first and last usability test task simple. 8. The **sample usability test report** includes a sample set of usability test tasks.
Usability test task scenario	A description of a **usability test task** phrased as a story or situation that test participants can put themselves in during a **usability test**. Note: 1. The difference between a **usability test task** and a usability test task scenario is that the latter is contextualized – that is, it contains additional contextual information that justifies the task. Sample **usability test task** for a calendar program: "Use the calendar program to set a reminder on December 7th." An example of a similar usability test task scenario "December 14th is the birthday of one of your friends. Use the calendar program to notify you a week in advance so you'll remember to buy a present." 2. Ideally, a **scenario** presents a context for a whole set of **usability test tasks**.

Creation of usability test tasks	The process of writing and improving **usability test tasks**. Notes: 1. To create a good set of **usability test tasks** proceed as follows: a. At first, don't look at the **interactive system** when generating ideas for tasks. Looking at the **interactive system** may bias you and cause you to create only tasks that can be solved by the **interactive system**. b. Ask current and prospective **users**, the development team, friends and colleagues to suggest appropriate **key tasks**. c. Get buy-in from the product team by working with team members to create the task set. d. Ensure that key **user requirements** and the goals of the **usability test** as defined in the **usability test plan** are reflected in the set of **usability test tasks**. 2. To improve a set of **usability test tasks**, proceed as follows: a. Solve your own **usability test tasks** using the **interactive system**. Discard tasks that turn out to be unreasonable. Keep tasks that are reasonable even if they are unsolvable or difficult to solve on the **interactive system**. It is not a goal of task creation to only create tasks that are easy to solve with the **interactive system**. b. Run a **pilot usability test session** and solicit feedback on the task set from the **test participant**. c. During the first 2-3 regular **usability test sessions** solicit feedback on each task by asking the **test participant** immediately after you hand them the task: - Is this task realistic? - Is this a task that you do in real life? d. During the **post-session interview** after the first 2-3 regular **usability test sessions** solicit feedback on the task set by asking the **test participant**: - What tasks are missing? e. If needed, modify the **usability test tasks**. Avoid sweeping changes. Make the smallest possible changes to the set of **usability test tasks** that will solve the problems.

Tips	**Types of DESIRABLE usability test tasks**
Open-ended	A **usability test task** that leaves it up to the **test participant** to define parts of the **goal** for the task. Examples: 1. "Have you been travelling recently or are you planning a trip? Where did you go or where are you planning to go? Please rent a car at your destination that suits your taste and purse for a period that would fit your stay." 2. "Buy whatever you want at this online music store for up to $50. We will reimburse you for all your expenses." Notes: 1. Open-ended tasks are useful because they motivate **test participants**. 2. Before a **test participant** starts solving an open-ended **usability test task**, they must explain the actual **goal** to the **moderator**. Otherwise the **moderator** may have problems understanding what the **test participant** is doing. 3. The **usability test report** must describe the **usability test task** that each **test participant** actually decided to carry out in order to help the reader understand what the **test participant** did. 4. Open-ended tasks may not work for a **usability test** of **prototypes** because a **prototype** usually will not support a task with user-defined data. 5. The **findings** from open-ended **usability test tasks** are more difficult to analyze than the **findings** from **closed** tasks, because open-ended tasks vary within a given frame.

Tips	Types of DESIRABLE usability test tasks
Closed	A **usability test task** that defines the **goal** of the task in detail. Examples: 1. "Rent a car from Avis.com. Pick up and return at London Heathrow Airport, Terminal 3. Pick-up time 10.30 am on Saturday 10 May. Return around 4.00 pm on Monday 12 May. Car class intermediate. Accept the standard CDW-insurance." 2. "Buy Tchaikovsky's Nutcracker Suite with Sergiu Celibidache published by Emi Classics." Notes: 1. A closed task is the opposite of an **open-ended** task. 2. Both **open-ended** and closed tasks are useful in **usability tests**. 3. In example 1, "Car class intermediate" and "CDW insurance" are undesirable **clues**. 4. In example 2, "Tchaikovsky" and "Sergiu Celibidache" are examples of undesirable **clues** if the task is handed out in writing.
Comparative	A **usability test task** that requires the **test participant** to compare two or more information items or procedures, often located on separate **interactive systems**. Examples: 1. "Find the cheapest microwave that can hold a standard size pizza." 2. "Find the microwave oven that best meets your needs." 3. "Register yourself at both Yahoo and Hotmail. Which registration procedure do you like best?" Notes: 1. Compare to task **contained within product**. 2. Comparative tests must be balanced because the first test object might influence the handling and judgment of the second.. Half of the **test participants** should first test A, then B. The other half should first test B, then A.
Contained within product	A **usability test task** that requires the **test participant** to locate information about one information item. Examples: 1. "What does the Samsung microwave M1753 cost?" 2. "Would the Samsung M1753 microwave oven meet your needs?" Note: 1. Compare to **comparative** task.
Subjective	A **usability test task** where the correct result depends on the subjective judgement of the **test participant**. Examples: 1. "Would the Samsung M1753 microwave oven meet your needs?" 2. "Find the microwave oven that best meets your needs." Note: 1. Compare to **objective** task.
Objective	A **usability test task** where the correct result is the same for all **test participants**. Examples: 1. "What does the Samsung microwave M1753 cost?" 2. "Find the cheapest microwave that can hold a standard size pizza." Note: 1. Compare to **subjective** task.

Avoid	Properties of usability test tasks that should be AVOIDED
Clues	Information in a **usability test task** that is unknown to most **test participants** and inadvertently helps them to solve the task. Example: 1. "Use Hotmail to create a personal signature. Having done so, send a short greeting to me (the moderator) using your personal signature." 2. "Find a song with Liza Minnelli". 3. "Use the help system to find information about CDW insurance". Notes: 1. The **usability test task** description in example 1 contains a clue: ”Signature” is a term used by Hotmail. It may not be known or understood by the **test participant**. The task tests the **test participants**' ability to recognize a keyword rather than the ability to understand the task. A similar task without clues is “You want your name and address to appear at the end of all messages you send. Show me a way to do this.” 2. It is generally recommended to hand out each **usability test task** in writing on a separate sheet of paper. However, certain types of tasks like the one shown in example 2 above contain significant clues if they are given in writing. One of the purposes of this task is to check the **error tolerance** of the search engine, since experience shows that few people are able to spell Liza Minnelli's name correct. Instead, phrase the **task** “Find a song by the artist the moderator will tell you”. 3. Both examples in the tip **Closed** task also contain clues.
Pretender	A **usability test task** that asks the **test participant** to pretend to be someone that they are not. Pretending to be someone you are not creates an invalid context for the task solution and may even insult the **test participant**. Example: 1. "You are a great admirer of Michael Jackson. What’s the name of his latest album?" 2. “You are a Spanish speaker interested in renting a car in Los Angeles (flying into LAX). You speak limited English. Find out how much it would cost to rent an intermediate size car for two days. “ Notes: 1. The task in example 1 is both a pretender task and an **unrealistic** task. A great admirer of Michael Jackson would hardly have to look up the name of his latest album. 2. Pretension within reasonable limits is OK, for example ”Imagine that you have a friend in Hamburg. You just arrived in Frankfurt Airport. Find the fastest train connection to Hamburg.”
Silly	A **usability test task** that attempts to entertain the **test participant**, often by using "humorous" names of persons, places or products. Example: 1. "Report noise damage caused by Mr. Makin A. Racket, 14 Pandemonium Avenue, 9999 Uproarton" (test of an insurance company’s website) Notes: 1. Silly tasks should be avoided because **test participants**' sense of humor often differ from the **moderator**'s. 2. Also, silly tasks may hurt the **test participants**’ trust in the **moderator** and cause them to be less frank in their answers. 3. Whenever your **usability test task** description distracts your **test participant** from the real task, you have failed – even if some **test participants** appreciate your humor.
Unrealistic	A **usability test task** that asks a **test participant** to do something that is unrealistic Example: 1. "Cancel the reservation" – This task is unrealistic if the **test participant** has not yet successfully completed a reservation.

Avoid	Properties of usability test tasks that should be AVOIDED
Derogative	A **usability test task** that asks the **test participant** to pretend that he or she sometimes behaves foolishly. Example: 1. "Let's assume that by mistake you have put two identical CDs into your shopping bag. Please get rid of one of them." Note: 1. Compare to **pretender** task
System-oriented	A **usability test task** that is not relevant from the **test participants**' point of view. Examples: 1. "Register yourself on this website" Instead, use a usability test task that requires registration. 2. "Use the help-system to find out what the date format is in this system" Instead, use a **usability test task** that requires entering dates to find out if there's a problem. If there are **usability problems**, note where **test participants** look for a solution. Note: 1. System oriented tasks are "necessary evils" from **users**' point of view. They are often subtasks of a relevant **task**. Instead, use a relevant **usability test task** that includes the system oriented task.
Sensitive personal information	A **usability test task** that encourages or forces the **test participant** to reveal sensitive personal information. Note: 1. Examples of sensitive personal information are a. Private address, e-mail address and telephone number b. Names, addresses, e-mail addresses and telephone numbers of friends or relatives c. Social security number d. Bank account number or credit card number e. The contents of the test participant's private email inbox.
Imprecise, hazy	A **usability test task** whose goal is unclear, so it is difficult to determine when the task is completed. Example: 1. "Find and read the cancellation policy" is imprecise, while the similar task "How much does it cost to cancel a reservation" is sufficiently precise.

Tips	Tasks in general
Key tasks	Test key **tasks** before testing specialized **tasks**. Example: 1. On an email website like Hotmail, **usability test** key tasks like "send email", "send email with attachment", "move email to folder", etc., before you **usability test** specialized **tasks** like "send me a notification on June 14".
Experienced users	Adapt your set of **usability test tasks** to the **test participants** in accordance with the **goals** of the **usability test** as defined in the **usability test plan**. Most often, sets of **usability test tasks** aim at novice **users**. If you have experienced **users** as **test participants**, use **usability test tasks** that reflect the needs of these **users**. Also consider having one or two **open-ended tasks** for these **test participants**, for example "Please show me what you did with this interactive system yesterday".

Tips	Tasks in general
Sunshine task	A usability test task that presumes that all equipment and the environment works perfectly. Note: 1. While sunshine tasks are basically OK, also include **test tasks** in your task set that test the **interactive system** under realistic, stressful conditions. Examples: 1. **Usability test** the behavior of the **interactive system** when the load on the system is non-trivial. For example, consider **usability testing** an email system with an inbox that is nearly full or contains 100.000 emails. 2. **Usability test** degradation behavior, for example how a smartphone application behaves when connectivity is limited.

3.2.6. Post-session interview – Debriefing

Term	Definition
Post-session interview	An activity in a **usability test session** where the **test participant** answers questions about his or her experience and general impression of the **usability** of the **interactive system**. Notes: 1. The purpose of the post-session interview is to let **test participants** share their experience in their own words without much prompting. 2. The post-session interview takes place after the **test participant** has carried out as many **usability test tasks** as time allows. 3. Recommended post-session interview questions: a. "Which 2-3 things did you like most about the system?" b. "Which 2-3 things are most in need of improvement?" c. "Would you buy this product? If yes, how much would you be willing to pay for it?" Ask this question only if it's relevant. d. "How likely are you to recommend this product to a friend or colleague? What makes you think so?" These questions are about opinions. They are helpful for prioritizing the **findings** from the **usability test**. 4. After the first 2-3 regular **usability test sessions** also solicit feedback on the task set by asking the **test participant**: - What tasks are missing? 5. Make the post-session interview as brief as possible. The standard questions should not take more than 2-3 minutes. Often, the **test participant** will have little to add to what they already said during the **usability test session**. 6. If **stakeholders** are observing the **usability test** in real time, invite them to ask questions to the **test participant**. Interacting directly with the **test participant** may help **stakeholders** get answers to questions that require special knowledge and may help convince them that the **test participant** is a valid representative of the **user group**. Stay along while **stakeholders ask** questions. Intervene if they start a. demonstrating all the wonderful things that the **interactive system** can do, b. directly or indirectly blame the **test participant**. 7. A post-session interview is also referred to as a debriefing.

3.3. Communicate findings

LO #	Learning Objective
3.3.1	Understanding the causes for communication problems and how to prevent them (K2)
3.3.2	Understanding the techniques for selling findings to stakeholders, for example the KJ-method, and the underlying principles (K2)
3.3.3	Mastering the analysis of findings (K3)
3.3.4	Knowing ways of communicating findings (K1)
3.3.5	Mastering the usability test report (K3)
3.3.6	Understanding positive findings (K2)
3.3.7	Knowing the video summary (K1)

Term	Definition
Finding	A result from a **usability evaluation**. Notes: 1. A finding can describe a. A **usability problem**. b. Something that **users** liked – that is, a **positive finding**. 2. A **classification and rating** should be assigned to each finding. 3. Findings from a **usability test** should be based solely on what **test participants** are able to accomplish with the **interactive system** when they carry out representative **usability test tasks**. Findings should not include opinions, neither from the **moderator** nor from the **test participant**. Personal opinions regarding potential **usability problems** are valid results from **inspections**. Examples: 1. Remarks like "I can do this easily but most others will have serious problems" from a **test participant** is a personal opinion rather than a finding. The **moderator** may obtain additional, valuable insight by following up with the question: "Why do you think so?" 2. It's OK to report findings that are based on opinions about an **interactive system**, for example "The design of the home page is really pretty", if they are voiced spontaneously by a substantial number of **test participants**.
Usability problem	A difficulty in using the **user interface** that affects the ability of the **user** to achieve their **goals effectively**, or **efficiently**, or with **satisfaction**. Note: 1. Usability problems can lead to confusion, error, delay, or outright failure to complete some **task** on the part of the **user**.
Positive finding	Something that **test participants** liked about the **user experience** or that clearly helped **test participants** achieve their **goals**. Notes: 1. At least 25% of the reported **findings** should be positive. If at first the **moderator** can only think of few things that **test participants** liked, the **moderator** should consider the **user interface** carefully. Often, positive features like reasonable response times and good support of **key tasks** are taken for granted and not explicitly acknowledged. 2. Reporting positive findings a. ensures that features that **test participants** liked are not removed simply because the development team was not aware that **test participants** appreciated them. b. creates a more positive attitude towards the usability study. 3. Positive findings must be completely positive – just as descriptions of **usability problems** should not contain positive elements. If a positive finding is not completely positive, separate it into a positive finding and a **usability problem**.

Communica-tion of findings	The process of informing **stakeholders** and others about the **findings** from a **usability test**. Notes: 1. The following techniques are available for this process: a. Informal discussions of **findings** with **stakeholders**, for example between **usability test sessions** b. Workshops with **stakeholders** c. Presentations of **findings** with subsequent discussion d. Writing and distributing a **usability test report** with subsequent discussion e. Creating and distributing a **video summary** f. Enter the **findings** in the bug database for the **interactive system**. **Communicators** should understand the rules that are in effect for entering bugs and try to follow them as carefully as possible. Some **usability problems** are quite different from bugs and need to be dealt with by **usability professionals** rather than developers. 2. Communication must be two-way. Never just present or distribute usability **findings**. Give directly affected colleagues, for example developers and designers, a chance to comment on the **findings** before they are distributed to others, and correct any misunderstandings. After distributing the **findings**, discuss them with **stakeholders**. 3. In order to decide on appropriate communication techniques, the **usability maturity** of the organization should be considered. a. If **usability maturity** is low, use appropriate techniques for **selling findings** to **stakeholders** by involving them in the planning and execution of the **usability test**. Also involve the **stakeholders** in determining the **findings**. b. If **usability maturity** is high, use the most efficient techniques, which are 1d and 1f. 4. For workshops with **stakeholders** consider the **KJ-method** or the principles that underlie it. 4. Regardless of which technique is used, listen carefully to feedback from the **stakeholders**. Avoid one-way communication or superficial dismissal of objections to the **findings**.
Selling findings	The process of convincing people who are not **usability professionals**, in particular **stakeholders**, that **findings** from a **usability test** should be taken seriously and acted upon. Notes: The following notes apply in particular for organizations with low **usability maturity** 1. **Usability testers**, in particular **communicators,** must understand **stakeholders** and their focus. If **stakeholders** do not accept the **findings** from a **usability test**, chances are that little beneficial change to the **interactive system**'s **user interface** will occur. 2. Some **stakeholders** are skeptical towards **usability** and **usability tests** as expressed by the anonymous quote “Cowboy programmers don’t need no stinkin’ usability.” Some developers view their **user interface** as an extension of themselves and thus take it personally when someone finds fault with it. 3. Knowledge of and agility in company politics is important for the success of **usability** in an organization. It's also important to understand the **usability maturity** of an organization. In an organization that has a low **usability maturity**, the best you can do is to make the consequences of ignoring **usability** visible through **usability testing** 4. Avoid opinions. There is no magic answer to the question “Why are your opinions better than mine?”, which easily leads to an "opinion war". Opinion wars only have losers. Be careful with opinion based methods, such as **inspection** and **heuristic evaluation**. **Stakeholders**, in particular developers, are skilled in discussing opinions. **Usability professionals** become new and interesting players if they "sell their ignorance" by insisting that they have no opinions and that only a **usability test** has the right answers. 5. Involve **stakeholders** in the planning and execution of the **usability test**. See the article “Involve stakeholders”.

3.3.1. Analyze findings

Term	Definition
Analysis of findings	The process that extracts **findings** from observations during **usability test sessions**. Notes: 1. Proceed as follows: a. During the **usability test session,** the **note-taker** records **usability** observations, usually by writing them down. Usability observations reflect events that cause problems with or have a positive effect on **effectiveness**, **efficiency** and **satisfaction**. b. After each **usability test session**, the **note-taker**, the **moderator** and the **host** meet to discuss the observations from the **usability test session** while everyone still remembers what happened. c. After all **usability test sessions** have been completed, the **moderator** and the **note-taker** separately extract 20-30 **usability findings** and 5-10 **positive findings** each from their observations. These **findings** reflect the observations that they consider most important. d. The **moderator**, the **note-taker** and the **communicator** meet and have a frank discussion about their **findings**. The **findings** are merged into a common list consisting of 20-30 **usability problems** and 5-10 **positive findings**. During this process, the approach described in **KJ-session** may be helpful, in particular if there are many different findings. The above figures ("20-30 **usability problems**" and "5-10 **positive findings**") are rules of thumb. However, it is important that a **usability test report** is usable. This means that the number of reported **findings** must be limited. For example, if you find 75 **usability problems**, you can't just report all of them and leave it up to the **stakeholders** to sort things out. It is a critical – and sometimes inconvenient – task for the **communicator** to prioritize the **findings** and report only the ones that are most important from a **usability** point of view. 2. It's crucial for the success of the analysis that the discussions between the **moderator**, the **note-taker** and the **communicator** are honest and based on observations rather than personal opinions. Bargaining during analysis should be avoided ("I'll accept this finding of yours without further discussion if you accept this finding of mine"). 3. For discussions of **findings** with **stakeholders** use the **KJ-method**. Lists of **usability problems** produced by the **moderator** and the **note-taker** as described in note 1 above should not be presented at the **KJ-session** because they may spoil the session by giving the **stakeholders** the impression that the **findings** have already been agreed. 4. Compare the **findings** from the current **usability test** to any relevant previous **usability test reports**. 5. **Findings** should be based on similar observations of at least two **test participants**. a. If just one **test participant** encounters a problem and all other **test participants** have few or no difficulties with the same issue, do not report a problem. b. If just one test participant encounters a serious or critical problem, and few other test participants have visited the page or window where the problem occurs, the **moderator** should use their experience, experience from others, and **heuristics** to judge whether the problem should be reported or not.

3.3.2. Usability test report

Term	Definition
Usability test report	A document that describes the findings from a **usability test**. Notes: 1. The purpose of the usability test report is to document and communicate the most important **findings** from a **usability test**. The report must be **effective** and **efficient** for the key **stakeholders**, in particular the development team and managers who make decisions about what will be changed. 2. A usability test report must be **efficient**. Specifically, this means a. Short and succinct. Report at most 50 **findings** regardless of the size of the **usability test**. The recommended number of **findings** is 5-25. See the tip **usability test report size**. b. Summary. Provide a summary for readers who are in a hurry, for example managers. The summary must be at most one page. The summary must be easy to find; this is usually accomplished by placing it at the very start of the usability test report, even before the table of contents. c. Uniform. All usability test reports from an organization should have the same structure and graphic layout. d. Comprehensible. Avoid usability jargon like "mental model" and "WCAG" (Web Content Accessibility Guidelines). 3. A usability test report should contain the following sections: a. **Executive Summary** b. Table of contents c. **Findings and recommendations** d. Description of the object of evaluation e. Purpose of the **evaluation**, including listings of or references to relevant **user requirements** f. **Evaluation method**, including **usability test script** and **list of test participants** Section d through f may be placed in one or more appendices. 4. A usability test report is always required for a **usability test**. However, the rules in note 3 are quite flexible. A simple, conforming report might be in Microsoft PowerPoint format with 1 slide executive summary, 2-3 slides describing the 5 most important findings, and 3 slides describing the object of **evaluation**, the purpose of the **evaluation**, and the evaluation method, respectively. 5. The following information should normally not be included in a usability test report: a. An explanation of what **usability** is b. A description of the **human-centred design** process c. Details about **recruitment**, for example copies of the **recruitment screener** and the **confirmation** sent to **test participants** d. Transcripts – that is, extensive verbatim text of what one or more **test participants** said during **usability test sessions**. However, short, verbatim text of what a **test participant** said such as quotes of 1-2 lines are recommended 6. A usability test report is also referred to as a "test report". 7. A free, **sample usability test report** that illustrates the requirements in this curriculum is available.

Term	Definition
Executive summary	A section in a **usability test report** that provides a brief overview of what was **usability tested** and the most important **findings** from the **usability test**. Notes: 1. The target audience for the executive summary is managers who need a brief overview of the most important **findings** from the **usability test**. 2. The executive summary should not exceed one page. 3. The executive summary should contain a. A brief description of the object of **evaluation** and b. When the test was conducted c. A brief description of the purpose of the **evaluation** d. A brief description of the **evaluation** method e. The 2-4 most important **positive findings** f. The 2-4 most important **usability problems** g. General **recommendations** based on the **findings** (optional). Examples: A procedure for reviewing the conformance of error messages to a set of agreed guidelines, or a thorough review of all UI texts that have been translated.
Findings and recommend-ations (section in usability test report)	A section in a **usability test report** that describes the 5-50 most important **findings** from the **usability test** and associated **recommendations for improvement** of the **interactive system**. Note: 1. The description of each finding should include a. **Classification and rating of finding** b. A header that briefly describes the finding c. An indication of the number of test participants who encountered a problem, for example “a few”, “most” or “all” test participants. d. A description of the finding e. Relevant quotes from test participants during their encounter of the finding (optional) f. **Recommendations for improvement** (optional) g. Screenshots illustrating the finding (optional)
Evaluation method (section in usability test report)	A section in a **usability test report** that describes how the **usability test** was conducted. Note: 1. The section "Evaluation method" should include a. Description of the evaluation design, for example the type of **evaluation** performed (**usability lab**, **remote usability test**, **unmoderated usability test**, **discount usability test**, etc.) and the experimental design of the **evaluation** b. Information regarding the physical and technical environment in which the **usability test** took place c. The **usability test script** d. An anonymized **list of test participants** e. Name and email address of the **moderator** or moderators who conducted the **usability test**.

Term	Definition
Classification and rating of finding	A measure given to a **usability problem** from a **usability test** to indicate the type of the finding, its impact and criticality on the **user experience** and the consequences. Notes: 1. The **moderator**, the **note-taker** and the **communicator** rate **usability problems** from the **test participants**' point of view. Sometimes, the ratings are done in cooperation with a domain expert. 2. Possible classifications are a. **Usability problem**. Each **usability problem** must have a severity rating as described in the following note b. **Positive finding** – Works well. This approach is recommendable c. Good idea - A suggestion from a **test participant** that could lead to a significant improvement of the **user experience**. d. Functional problem - Bug 3. Typical severity ratings of **usability problems** are: a. Minor: Minor dissatisfaction; noticeable delays; or superficial difficulties. b. Major: Substantial delays; or moderate dissatisfaction c. Critical: **Test participants** gave up – showstopper; substantial dissatisfaction; or minor financial damage to **user**. d. Catastrophic: Existential threat. Potentially life-threatening; or bodily harm; or substantial financial damage. 4. Important parameters that influence severity ratings are a. Frequency – How often does the **usability problem** occur? b. Impact – How badly does it hit the **user** and the **user's** environment when it occurs? c. Persistence – How quickly will **users** learn to avoid the **usability problem**? 5. The above definitions of severity ratings are used in practice but at this time, unfortunately, there is no set of generally agreed severity ratings. 6. Avoid severity ratings like "Imperative to fix as soon as possible" or "Must fix, regardless". **Usability testers** are responsible solely for **usability**, not for cost to fix, etc. Comments of this type overstep the charter of the **usability professional**.

Term	Definition
Recommen-dation for improvement	A suggestion for how a **usability problem** can be solved. Notes: 1. Recommendations are optional in **usability test reports**. 2. Emphasize that recommendations are just one of several ways to solve a **usability problem**. Start your recommendations by stating for example "One way of solving this problem is to ..." 3. Arguments for including recommendations in a **usability test report**: a. Recommendations help **stakeholders** understand what the **usability problem** really is. In other words: The recommendation extends the problem description. b. Many **stakeholders** expect such advice from **usability professionals**: "That's why we hired you!" c. Missing recommendations create a vacuum. This vacuum is often filled with unusable solutions. 4. Arguments against including recommendations: a. The solution is obvious or the **communicator** knows the **stakeholders** well enough to trust that they don't really need help – or, worst case, might actually find it insulting. b. Unfamiliarity with all the constraints (technical, business, legal, political, etc.) c. Lack of confidence that the recommendation is a good one. d. The problem is large enough in scope that the **communicator** can't easily describe a solution in a few sentences. Example: Major problems understanding the navigation structure. Complex problems are best solved by team collaboration. Collaboration also increases buy-in, that is, the chance that the recommendations will actually be implemented. 5. Means for expressing recommendations: a. Text b. Altered screenshots, possibly with callouts 6. A great recommendation is **effective** and **efficient**. It: a. solves the problem b. recommends the least possible change which leads to the intended outcome c. is illustrated, for example with an altered screenshot d. is justified e. provides detail f. speaks the readers' language g. is constructive and direct h. is short i. addresses only the original **usability problem**.

Tips	Usability test report
Document findings	Always document the findings of a **usability test** – even if you use the **KJ-method**, and even if you have little time, and even if you use **agile usability evaluation**.
Usability test report size	The quality of a **usability test report** is not proportional to the number of reported **findings**. On the contrary, it is important that **communicators** prioritize **findings** and only report the essential **findings**, independent of the extent of the **usability test**. There is an important exception to this rule of thumb: All catastrophic and critical **usability problems** must be reported, even if this means that the total number of reported **findings** exceeds 50. Of course, this happens rarely.

Tips	Usability test report
Sample usability test report	A CPUX-UT Usability Test Report Example produced by the UXQB is available free of charge from www.UXQB.org in an English and a German version. The Usability Test Report Example illustrates many of the definitions given in this curriculum. It includes 1. A sample executive summary 2. Sample **findings**. 3. A sample **usability test script**, including **usability test tasks,** The Usability Test Report Example includes rules and tips for writing good **usability test reports** that are not contained in this curriculum. These additional rules and tips are not part of the knowledge required to obtain a CPUX-UT certificate.
List of test participants	1. Provide key characteristics of **test participants** that are relevant to the validity of the **usability test**. This could include a. Profession or job title b. Previous knowledge of and interest in the **interactive system** being tested c. Previous knowledge of and interest in similar **interactive systems** d. Age or age group e. Sex (note that sex is often of less importance) 2. Do not provide names of **test participants** – not even surnames – or any other information that could identify a **test participant** 3. Provide meaningful information about profession or job title. For example, "manager", "student" or "retired" are not helpful, while "HR manager", "graduate student, organic chemistry" and "butcher, retired" are helpful.
Examples in general findings	Include examples in general **findings**. Example: 1. **Findings** like "The website uses jargon which was unfamiliar to test participants" are unusable because they are expressed broadly. They should be accompanied by at least 2 specific examples.
Tactful findings	**Findings** should be expressed tactfully. Rude **findings** create enemies – not more usable **interactive systems**. Examples of rude **findings**: 1. "The availability of the books is not shown!!!" 2. "The design of the home page is unprofessional and sloppy."
One issue per finding	**Findings** must not be conglomerates of several independent but possibly related **usability problems**.
Icons for ratings	Use icons to show **severity ratings** in **usability test reports**. Use icons that are intuitive and have considerable contrast so readers can easily locate for example all **usability problems** that have the **ratings** critical or catastrophic. See the **sample usability test report** for examples.

3.3.3. The KJ-method, affinity diagramming

Term	Definition
KJ-method	A brainstorm-based method for quickly establishing consensus among **stakeholders** regarding the most important **findings** from a **usability test**. Notes: 1. For a more detailed description of the actual approach, see **KJ-session** 2. The KJ-method is useful for a. **Selling findings** to **stakeholders**, particularly in organizations with a low **usability maturity**. b. **Analysis of findings** 3. The KJ-method is essentially a brainstorm, in particular the first part (**KJ-session** step 1c through 1e). In order not to inhibit the brainstorm, discussions are not allowed during this part. 4. Several other methods for consensus building exist, for example **affinity diagramming**. 5. The method is named after the Japanese ethnologist Jiro Kawakita.
KJ-session	A workshop lasting about 1 hour where the **KJ-method** is applied Notes: 1. A KJ-session consists of the following steps: a. Encourage all **stakeholders** to observe one or more **usability test sessions** and take notes. b. Invite all **stakeholders** who have observed at least one **usability test session**. Conduct the KJ-session immediately after the final **usability test session**. c. Ask each participant to write down the most important **usability problems** they have observed during the **usability test**. Each **usability problem** is written on a separate card or post-it note. d. Display the cards, for example on a wall or a blackboard. Self-adhesive post-it notes are useful for this purpose. e. Ask participants to study each other's cards. If a card inspires a participant to think of additional important **usability problems**, they can write them on additional cards and add them to the displayed cards. f. Sort the cards into groups and combine descriptions of the same **usability problem**. Combine **usability problems** only if there is full agreement among the participants; if just one participant objects, refrain from combining the **usability problems**. g. Name each group. Use cards of a different color for group names. h. Vote for the most important **usability problems**. Each participant should place 10 marks on the **usability problems** that they consider most important. For example, a participant may place two marks on each of 5 **usability problems**, or they may place 10 marks on one particular **usability problem**, or any other combination. Source www.uie.com/articles/kj_technique/ 2. **Positive findings** are not considered in a KJ-session. 3. The optimal number of active participants in a **KJ-session** is 3-8. 4. The **communicator** organizes and moderates the KJ-session. In addition, the **moderator** and the **note-taker** should attend the KJ-session. The main role of the **communicator** is to make the KJ-session happen and to make it run smoothly. The **communicator** should explain how the method works, observe and take notes. Depending on the context and the attitude of the **stakeholders**, the **usability testers** should be passive or participate actively by adding cards, commenting on cards, participating in the sort, and voting. Note that the method works well even if the **usability testers** do not participate actively, and passive participation can make the results more convincing to the **stakeholders**. 5. **Stakeholders** who have not observed at least one **usability test session** must remain passive because their contributions are opinions.

Affinity diagramming	A hierarchical technique for organizing and grouping the issues and insights across large quantities of qualitative data and showing it in a visual display, usually posted on the wall of a room. Note: 1. When used to analyze findings, affinity diagramming is similar to the **KJ-method**.

Tips	KJ-method
Main strengths	The main strengths of the **KJ-method** are: 1. The **findings** are defined by the people who will implement them – that is, the **stakeholders**. This increases the acceptance of the **findings**. 2. The **findings** are available immediately after the **KJ-session**. The product team can start to correct **usability problems** immediately. There is no need to wait for a formal **usability test report**. 3. Everyone gets heard.
Group names	Use the group names decided during the **KJ-session** as section headings in the **usability test report**.
Combining problems	Combining **usability problems** is easy if you put comments on self-adhesive post-it notes. Combine **usability problems** by sticking them on top of each other.
The role of the usability professional	Some **usability professionals** don't like the **KJ-method** because they believe that they are the only ones who can analyze and describe the **findings** of a **usability test**. Experience shows, however, that **stakeholders** most often arrive at **findings** that are similar to what the **usability professionals** have arrived at.

3.3.4. Video summary

Term	Definition
Video summary	A video that illustrates some of the most important **findings** from a **usability test** through appropriate video clips from **usability test sessions**. Notes: 1. The purpose of the video summary is to briefly present the most important **findings** from a **usability test** through video clips showing **test participants** struggling with the **interactive system**, or solving a task effortlessly. The target group for the video summary is key **stakeholders** who were unable to observe **usability test sessions** live. 2. The length of a video summary should be 3-15 minutes. 3. Include at least one **positive finding** in the video summary. 4. Illustrate each **usability problem** or **positive finding** with at least 2 clips where different **test participants** experience the **finding** to show that more than one **test participant** encountered the **usability problem**. 5. Start the video by displaying a title for so long that viewers have time to read the essential information in the title (suggestion: 8 seconds): a. Object of evaluation, b. Test date, c. Name of the **communicator** who created the video summary d. Where to find the **usability test report**, for example an intranet link. 6. Use subtitles or title slides at the start of the clips for a **usability problem** or **positive finding** to briefly explain what the point is. 7. Avoid "talking heads" – that is, videos of the **communicator** explaining what the **finding** is. 8. Appropriate video clips can also be integrated into the **usability test report**. 9. It takes considerable time to create a good video summary and appropriate video clips, for example one person hour per minute of the video summary. Create video summaries and video clips only if they pay off – that is, if they are actually viewed by **stakeholders** and cause usability improvements.

3.4. Roles in a usability test

This section describes the **roles** of the key actors in a **usability test**.

LO #	Learning Objective
3.4.1	Mastering the moderator, note-taker and communicator roles (K3)
3.4.2	Understanding the observer, host, administrator and test participant roles (K2)

Term	Definition
Role	A duty or function that a person carries out within an organization. Notes: 1. A role describes a set of connected behaviors, rights, obligations, and norms in a job situation. 2. Roles are occupied by individuals, who are called actors. 3. Depending on the complexity of a project, several people might share one role or multiple roles might be assigned to one person. 4. Work roles are "hats" that people wear when they take on the corresponding job responsibilities and perform the associated activities. Example: 1. In a **usability test session**, one person can simultaneously occupy the roles **moderator** and **note-taker**. In another **usability test session**, in contrast, one person may act as **moderator** and two other people may both act as **note-takers**.
Usability tester	A **usability professional** who evaluates **user interfaces** in various stages of realization. Notes: 1. Usability tester is a generic **role** covering the **roles moderator, note-taker, host**, **communicator** and **evaluator** (a person who carries out an **inspection**).
Moderator	A neutral **usability professional** who conducts a **usability test session**. Notes: 1. The **moderator**'s responsibilities during a **usability test session** are described under **usability test session**. 2. Not everyone has the empathy and patience to be a good **moderator**. 3. The **moderator** must be curious and follow up on any unclarities and unexpected happenings. 4. Every **moderator** should try at least once to be a **test participant**. 5. Facilitator is sometimes used as a synonym for **moderator**.
Note-taker	A **usability professional** who makes notes of **findings** during a **usability test session**. Notes: 1. The use of a note-taker allows the **moderator** to fully concentrate on the **test participant**. 2. There is no generally accepted shorthand for note-taking during a **usability test**. Some note-takers prefer to print copies of screenshots and enter notes directly on the printed screenshots. Others prefer to speed-type on a laptop. However, the noise of the keyboard can be annoying and may constantly remind **test participants** that their actions are observed and recorded. Hand-written notes might be preferable. 3. The note-taker also plays an active role during the **analysis of findings**.
Observer	A person who watches **test participants** who carry out **usability test tasks** in a **usability test session**. Notes: 1. Observers are often **stakeholders**. 2. Observers do not interfere with the **moderation**. 3. Observers may be actively involved in the **analysis of findings**, for example when the **KJ-method** is used.

Term	Definition
Host	A **usability professional** who services **observers** during a **usability test session**. Notes: 1. Often, the **note-taker** also handles the host **role**. 2. Some tasks for the host are: a. Help **observers** understand what is going on during the **usability test session**. b. Encourage **observers** to take notes during a **usability test session**. c. Facilitate discussions between **observers**. d. Take notes from discussions between **observers** to preserve important insights. e. Participate in the discussion of observations in the **analysis of findings** after each **usability test session**, in particular to convey what **observers** said during the session.
Communi-cator	A **usability professional** who communicates **findings** from a **usability test**, for example by moderating a **KJ-session** or by writing and presenting the **usability test report**. Note: 1. The communicator also plays an active role during the **analysis of findings**.
Administrator	A person who administers **usability tests**. Note: 1. Administrative tasks include scheduling **test sessions**, **test participant** management, communication with **test participants** before and after the **usability test session**, and administering **incentives**.
Test participant	A representative **user** who solves typical **usability test tasks** in a **usability test session**.

3.5. Quantitative usability test

LO #	Learning Objective
3.5.1	Understanding the quantitative usability test (K2)
3.5.2	Understanding the parameters that are most often measured in a quantitative usability test (K2)
3.5.3	Understanding the purpose of confidence intervals in relation to quantitative usability test measurements (K2)
3.5.4	Understanding the number of test participants in a quantitative usability test (K2)
3.5.5	Understanding failure rate and disaster rate (K2)
3.5.6	Understanding invalid measurements and how to handle them (K2)
3.5.7	Understanding outliers (K2)

Term	Definition
Quantitative usability test	A **usability evaluation** that focuses on obtaining figures for the **effectiveness**, **efficiency** or **satisfaction** of an **interactive system**. Notes: 1. Quantitative usability tests are used to a. Evaluate a design against **quantitative user requirements**. The evaluation process is described in the CPUX-F curriculum. b. Compare **interactive systems**, for example competitive products c. Compare versions of an **interactive system**. 2. The parameters that are most often measured in quantitative usability tests are a. **Task completion time** - that is, time to complete a **usability test task** (a measure of **efficiency**) b. **Success rate** (a measure of **effectiveness**) c. **Failure rate** (a measure of **effectiveness**) d. **Disaster rate** (a measure of **effectiveness**) e. **Satisfaction** **Success rate** + **Failure rate** = 100% 3. To plan and analyze measurements from a **quantitative usability test**, some knowledge of the following basic statistical concepts is required: a. **Mean** b. **Geometric mean** c. **Standard deviation** d. **Confidence interval** d. **Outlier**. If you are in doubt about these concepts, or if your quantitative usability test is the basis for important decisions, consult an expert on statistics. 4. Never report just time or mean time. Always include the **confidence interval (CI)** when reporting quantitative results.. There is a tremendous difference between the measurements 100, 110, 120 seconds and 10, 110, 210 seconds even though they have the same **mean**. 5. In order to obtain valid measurements it is of particular importance to recruit representative **test participants**. 6. An attempt to solve a **usability test task** during a quantitative usability test can have the following outcomes: Success, failure, disaster or **invalid measurement**. 7. Participants should be briefed that time is being measured, and that they should report when they feel theyhave solved a task, or give up. Add “This is not to put pressure on you – we just want to know how long things take.” 8. Assists like “Please read the task again” are permissible. More extensive assists like “The answer is not on this screen” are not permissible. 9. Qualitative **findings** discovered as part of a quantitative usability test should also be reported in order to help **stakeholders** determine how unsatisfactory performance can be improved. 10. In a quantitative usability test, **think aloud** should be discouraged as it may influence measurements. 11. **Satisfaction** is measured with a **user survey**, for example **SUS**. 12. Compare to **qualitative usability test**.

Term	Definition
Task completion time	End time minus start time for a **test participant**'s solution of a **usability test task**. Notes: 1. Start time: Timing should start after the user has read the instructions and at the point that information on the interface is presented to the user. Start time is not the time when the **test participant** receives the **usability test task**. 2. End time is the time when the **test participant** declares that he or she has found the correct answer to the **usability test task**, and the answer is indeed correct. The **test participant** does not have to read out the answer, but it's not sufficient that the answer appears on the screen. 3. The reported task completion time is the **geometric mean** value of all task completion times for all successful solutions of the **usability test task**. 4. If a **test participant** is unable to solve a **usability test task**, or if they end up with an incorrect result, then the task completion time is not considered. Instead, the task attempt is counted as a failure. 5. Determining the exact starting and ending time can be difficult. The uncertainty is often 5 or more seconds. Examples of uncertainty: a. The **test participant** starts working on the **usability test task**, but then returns to the instructions and reads them again carefully. b. The **test participant** forgets to say that they have completed the task. The correct answer is on the screen but it is unclear whether the **test participant** has perceived it.
Success rate	The percentage of all attempts to solve a **usability test task** that were successful. The success rate is a measure of the **effectiveness** of the **interactive system**. Note: 1. Success: A task attempt is a successful if the **test participant** manages to find an acceptable answer without any assistance from the **moderator** 2. Report both the success rate and the corresponding **confidence interval**.
Failure rate	The percentage of all attempts to solve a **usability test task** that are failures. Note: 1. Failure: A task attempt is a failure if the **test participant** a. Gives up b. Is unable to find an acceptable answer within a reasonable period of time A common time limit is 10 minutes. c. Finds an incorrect answer and believes it is correct. Compare to **Disaster rate.** 2. Failures include disasters. 3. Report both the failure rate and the corresponding **confidence interval**.
Disaster rate	The percentage of all attempts to solve a **usability test task** that are disastrous. Notes: 1. An attempt to solve a **usability test task** is disastrous a. if the **test** participant finds a seriously wrong answer and considers it correct - that is, would continue their work using the wrong answer b. the consequences of using the wrong answer are critical or an existential threat. 2. Note the important distinction between failure (as defined in **Failure rate**) and disaster. Usually, the number of disasters is much smaller than the number of failures. 3. Example of a disaster for a council website: **Test participant** finds wrong date for the next collection of bulky waste. 4. Example of a failure for a council website: **Test participant** finds an outdated telephone number for city services. This is not a disaster since **users** would easily find out that the number is wrong when they call it so the consequences from the **user**'s point of view would be limited.

Term	Definition
Invalid measurement	A **usability test task** measurement that cannot validly be included in the results of a **quantitative usability test** because of equipment malfunction, software malfunction, procedural error, **test participant** error, or irrelevant interruption. This list is not exhaustive. Notes: 1. Examples of reasons for declaring the measurement of a **usability test task** invalid: a. **Moderator** helps **test participant** with **usability test task** Note: Seeking and getting assistance from support or the help system does not invalidate a measurement, but the time used to contact support, waiting in a queue, etc., is included in the **task completion time** b. **Moderator** intervenes to solve a technical problem c. Hardware or software problems on the computer d. Required basic software not installed on computer For example appropriate Adobe Reader missing e. **Test participant** interrupted for more than 10 seconds For example by a colleague or a telephone call (unless the interruption is part of the **usability test task**) f. **Test participant** solves or starts solving a wrong **usability test task** g. **Test participant** misunderstands the **usability test task** or does not read it with sufficient care h. **Test participant** starts discussing with the **moderator** Invalidates the measurement if the discussion lasts more than 10 seconds.
Mean	The sum of all measurements divided by how many measurements there are. Notes: 1. The mean is often referred to as the average or the arithmetic mean. 2. Never report just a mean. Always include the **confidence interval**. Example: 1. The mean of 5, 6, 7, 7, 7, 8, 9, 9, 9, 23 minutes is 9 minutes
Geometric mean	A type of **mean** or average, which indicates the central tendency or typical value of a set of numbers by using the product of their values as opposed to the arithmetic mean which uses their sum. The geometric mean is defined as the nth root of the product of n numbers. Notes: 1. The geometric mean is often preferred over the **mean** because it is less sensitive to **outliers** like “23” in example 3 below. 2. The GEOMEAN function in Microsoft Excel computes the geometric mean. Examples: 1. The geometric mean of 2 and 8 is 4, because the square root of 2*8 is 4. 2. The geometric mean of 2, 10 and 11is 6.0 because the third root of 2*10*11 = 220 is 6.0 3. The geometric mean of 5, 6, 7, 7, 7, 8, 9, 9, 9, 23 minutes is 8.2 minutes.
Standard deviation	A measure of how spread out the measurements are. The standard deviation is defined as the square root of the variance. The variance is the average of the squared differences from the measurement to the **mean**. Notes: 1. The standard deviation is helpful in determining **outliers**. 2. The STDEV function in Microsoft Excel computes the standard deviation.

Term	Definition
Confidence interval	An estimate of a range of values that includes the true value. Notes: 1. The purpose of the confidence interval is to show the interval where the correct value lies with some certainty, for example 95%. This certainty is called the confidence level. The correct value is the value that would be obtained if all users performed the **usability test task**. 2. Reported measurements should always be accompanied by a confidence interval. In other words, do not just report that the time to solve a particular **usability test task** was 140 seconds. Instead report that the time to solve the **usability test task** is between 90 and 210 seconds with 90% certainty. 3. Increasing the number of representative **test participants** in a **quantitative usability test** decreases the length of the confidence interval. 4. To arrive at a reasonably small confidence interval, a minimum of 20 **test participants** is often required. It is not unusual to have **quantitative usability tests** that involve 100 **test participants** to arrive at an acceptably small confidence interval. 5. The computation methods for the confidence interval are different for **success/failure/disaster rates**, **task completion times**, and satisfaction scores. For additional information including calculators for computing CIs and more, see Jeff Sauro's website www.measuringusability.com
Outlier	A measurement that is more than twice the **standard deviation** away from the **mean** or falls below the minimum time an expert user requires to do the task. Note: 1. Check outliers carefully. Long-runners may be representative of a small but important group of careful users. They are characteristic of time distributions. Do not just discard them, because they may represent valid albeit extreme data. Examples: 1. The measurements 5, 6, 7, 7, 7, 8, 9, 9, 9, 10 minutes contain no outlier. 2. The measurements 5, 6, 7, 7, 7, 8, 9, 9, 9, 23 minutes contain the outlier “23”, because the **mean** is 9, and the **standard deviation** is 4.8, so any measurement greater than (9 + 2 * 4.8) = 18.6 is an outlier.

3.6. Variants of usability test

This section briefly describes a number of commonly used variants of **usability test**.

LO #	Learning Objective
3.6.1	Understanding discount usability test and how it differs from an ordinary qualitative usability test (K2)
3.6.2	Knowing remote usability test, unmoderated usability test, A/B testing and RITE (K1)
3.6.3	Knowing important differences between remote usability test, unmoderated usability test, A/B testing, RITE and ordinary qualitative usability test (K1)
3.6.4	Knowing retrospective recall (K1)
3.6.5	Knowing eye tracking (K1)
3.6.6	Knowing international testing (K1)

Term	Definition
Quantitative usability test	See section 3.5
Discount usability test	A **qualitative usability test** where the **usability tester** puts emphasis on keeping costs down without compromising too much on the **quality of the usability evaluation**. Notes: 1. A discount usability test has one or more of the following characteristics: a. The **moderator, note-taker, host, communicator** and **administrator** are one person. b. No **usability lab** is used; **usability test sessions** take place for example in a meeting room c. The test is limited to 5 **test participants** d. The **usability test report** is short and lists a limited number of **findings**, for example up to 15. Of course, this could mean that the **communicator** will need to prioritize the **findings** because not all **findings** can be reported. The goal is to write an **effective** and **efficient usability test report**, not a complete **usability test report**. e. The **usability test** takes less than 30 person hours including the **communication of findings**. 2. A discount usability test may include an **expert usability review**.
Remote usability test	A **usability test** where the **test participant** and the **moderator** are in different physical locations. Notes: 1. The **moderator** observes the **test participant** using an internet connection. 2. The **moderator** communicates with the **test participant** over the telephone or via an internet connection. 3. Compare to **usability test** and **unmoderated usability test**.

Term	Definition
Unmoderated usability test	A **usability test** where **test participants** solve **usability test tasks** without being observed. Notes: 1. The main advantage of an unmoderated usability test is that **recruitment** is fast and cheap. The analysis effort is the same as for a traditional **usability test**. 2. **Test participants**' actions are usually video recorded for later analysis. 3. Unmoderated usability tests are often conducted on the **test participant's** computer in their home. Video recording is accomplished through recording software installed on the computer by a vendor of unmoderated usability test services. 4. Unmoderated usability tests are offered by several vendors who will **recruit test participants** and conduct **usability test sessions** using your **usability test script** for a reasonable fee, currently $40 - $100 per **test participant** for 15 minutes of test time. 5. Sample vendors of unmoderated usability tests are usertesting.com, loop11 and userzoom. 6. Important quality parameters when selecting a vendor are clear voice recordings, good support, warranty, and a recording tool that does not interfere with the **interactive system** that is being tested. 7. Compare to **Usability test** and **Remote usability test**. 8. "Unmoderated usability test" is sometimes referred to as "Crowd usability test" even though unmoderated tests do not require a large number of test participants.
Retrospective recall	An activity that takes place after a **usability test session**. The **test participant** watches the video of the **usability test session** and comments on their deliberations and actions during the **usability test session**. Notes: 1. Watching and commenting on the video of the **usability test session** takes approximately as much time as running a new **usability test session** with a fresh **test participant**. Before using this method, consider carefully if the retrospective recall will provide sufficient insight to justify the time investment. 2. Retrospective recall may be required when you don't want the **test participant** to **think aloud** for one or more of the following reasons: a. You are doing a **quantitative usability test** b. The **test participant** is involved in tasks where distraction is potentially life-threatening, for example air traffic control, factory assembly lines or medical systems c. Tasks require so much concentration that **test participants** would be overtaxed by being asked to **think aloud** while working 3. Retrospective recall is not observation of actual **test participant** behavior. Retrospective recall may be influenced by irrelevant factors, such as post-rationalization and attempts to justify non-optimal behavior.
RITE (Rapid Iterative Testing and Evaluation)	A **qualitative usability test** where changes to the **user interface** are made as soon as a **usability problem** is identified and a solution is clear. Notes: 1. The RITE method focuses on instant redesign to fix problems and then confirming that the solution works with new **test participants**. 2. Changes can occur after observing as few as one **test participant**. Once the data for a **test participant** has been collected, the **usability tester** and the **stakeholders** decide if they will be making any changes prior to the next **test participant**. The changed **user interface** is then tested with the remaining **test participants**.

Term	Definition
Eye tracking	Tracking and recording **test participants'** eye movement during a **usability test**. Notes: 1. The purpose of eye-tracking is to get an understanding of where **test participants** look at the screen, and where they do not look. 2. Eye tracking often generates a heat map showing where **test participants** looked over a period of time, for example while solving part of a **usability test task**. 3. Eye tracking does not show **test participants'** thought activity. 4. Eye tracking rarely generates important insight that could not have been achieved through an ordinary **usability test**. However, findings from a **usability test** that includes eye tracking are often more easily accepted by **stakeholders**.
International usability test	A **usability test** that is conducted in several countries. Notes: 1. The purpose of an international usability test is to understand **users**' reactions to an **interactive system** outside of the country where it was built and **usability tested**. 2. An international usability test should be conducted with a native **moderator** and native **test participants** who communicate using their primary language during the **usability test sessions**. Reliability suffers if **usability test sessions** are conducted in English when English is not the primary language of the **test participant**. **Observers** who do not understand the local language will have to rely on simultaneous translation. The abilities of the translator are of great importance.
A/B testing	A way to test changes to the design of an **interactive system** against the current design and determine which changes produce positive results. Notes: 1. A and B can be competing designs and each version is served to half the actual **users** of the **interactive system**. Other times, A is the current design that most **users** see and B, which might be more daring, is served only to a small percentage of **users** until it has proven itself. 2. The purpose of A/B testing is to validate that a new design or change to an element in an **interactive system** is improving the **interactive system** before the production version is modified. 3. **Usability tests** give qualitative insight into **usability problems**. In contrast, A/B testing offers concrete numbers. With A/B testing the caveats for uncertainty in **quantitative usability tests**, in particular **confidence intervals**, apply. 4. A/B testing is also known as split testing.

3.7. Ethical rules for usability tests

LO #	Learning Objective
3.7.1	Understanding why ethical rules for usability tests are necessary (K2)
3.7.2	Understanding the ethical rules for usability tests (K2)

Term	Definition
Ethical rule	A statement that describes a commonly accepted norm for decent behavior The ethical rules that are most important for **usability testing** are: 1. Make sure that the **test participant** feels safe and important. a. At the end of the **usability test session** the **test participant** must feel at least as comfortable as at the start. b. Emphasize that we're never evaluating the **test participant** but the **interactive system**. c. Maintain a professional neutral demeanor. Don't indulge in sarcasm or flirt. 2. De-mystify the **usability test** by telling the **test participant** in advance what to expect. a. Avoid words like "lab", "test subject", "experiment". 3. A **test participant** can never make a mistake or do anything "stupid" or "wrong" during a **usability test session**. 4. **Test participants** can stop at any time and still get their **incentive**. 5. **Test participants** are anonymous. a. Never put any direct or indirect pressure on a **test participant** to reveal personal information such as their email inbox. b. If a **test participant** voluntarily decides to disclose personal information, that's OK. 6. Use an **informed consent agreement** to obtain informed consent for use of all data collected. Notes: 1. Ethical rules are necessary to ensure that **usability professionals** do no harm to other human beings, and that they respect privacy, confidentiality and anonymity. 2. The ethical rules in this curriculum are based on the User Experience Professionals' Association's Code of Professional Conduct (CPC), https://uxpa.org/resources/uxpa-code-professional-conduct 3. If you are a member of other professional associations, for example the American Psychological Association, additional or different ethical rules may apply.

Tips	Ethical rules
De-stressing	Try to soothe the **test participant** ahead of the **usability test session**. Offer a cup of coffee or a soft drink. Engage in small talk while you walk the **test participant** to the **test location**. A cup of coffee and a short break works wonders if a **test participant** gets stressed.
Quick success	The best way to make a **test participant** relax is by assuring that they have a quick success in solving the first **usability test task**.
"Funny" video clips	Showing video clips of "funny" episodes – such as **test participants** picking their nose while deliberating – to entertain visitors of your **usability lab** is unethical.
Inappropriate jokes	Jokes like "If you can't figure out how to fill out this form, you will not get your gift" are unethical.

4. User survey

LO #	Learning Objective
4.0.1	Understanding the user survey (K2)
4.0.2	Understanding the qualitative user survey (K2)
4.0.3	Understanding the quantitative user survey (K2)
4.0.4	Understanding the difference between qualitative and quantitative user surveys (K2)
4.0.5	Understanding the questionnaire (K2)
4.0.6	Understanding usability criteria for questionnaires (K2)

CPUX-UT – Curriculum

This curriculum breaks down a **user survey** of an **interactive system** into the following steps:

1. Define the goals of the **user survey** in cooperation with the **stakeholders**
   a. Write a survey plan (similar to a **usability test plan**) that describes the goals of the **user survey**, including questions that **stakeholders** want to have answered and **users' goals** for the underlying **interactive system**
   b. Ask **stakeholders** to review and approve the survey plan
2. Develop the **questionnaire**
   a. **Interview users** and **stakeholders** to determine questions in the **questionnaire**
   b. Write the **questionnaire** in accordance with the usability rules in the definition of **questionnaire**.
   c. Describe how responses will be analyzed to serve the **goals** of the **user survey**.
   d. Ask **users**, **stakeholders** and survey experts to review the **questionnaire** and the plan for analyzing reponses.
3. **Test** the **questionnaire** for clarity
   a. Test the questionnaire for clarity by asking 5 representative users to fill out the questionnaire one by one while they think aloud so the **moderator** can check that the **test participant**s' understanding of the **questionnaire** and the questions match the intentions.
   b. If the **questionnaire** doesn't work, correct the problems in the **questionnaire** and continue from step 3a.
   c. Send out the **questionnaire** to for example 25 prospective **user survey participants**.
      - Ask them to fill out the **questionnaire**
      - Ask for frank comments regarding the **questionnaire**.
      - Check the responses carefully.
   d. If it doesn't work, correct the problems and continue from step 3a.
   e. If it works, then the **questionnaire** is ready for deployment
4. Carry out the **user survey**
   a. Recruit **user survey participants**
   b. Send the **questionnaire** to the **user survey participants**
   c. Follow up with a reminder to the **user survey participants** who have not responded within a reasonable time, if possible
5. Analyze the responses from the **user survey**
   a. For a **qualitative user survey**, use the **KJ-method** or **affinity diagramming**
   b. For a **quantitative user survey**, use the advice regarding basic statistical concepts provided in section 3.5 about **quantitative usability test**
6. Communicate the results of the **user survey**
   a. Write a report that answers **stakeholders**' questions in a useful and usable way
   b. Present the results of the **user survey**

Term	Definition
User survey	An **evaluation** where **users** are asked to report subjective data into a **questionnaire** based on their experience in using an **interactive system**. Notes: 1. User surveys are used to evaluate **users' satisfaction** with an **interactive system** and to gather **context of use** information. 2. Possible user survey **goals** are: a. Objectifying **satisfaction** or dissatisfaction with an **interactive system** b. Measuring effects of changes in an interactive system with user surveys before and after the change c. Comparing two **interactive systems** 3. The process for creating a useful and usable user survey is described in the beginning of this section. 4. The number of **user survey participants** in a user survey can vary from a few, for example 5 **test participants** in a **usability test**, to thousands. 5. A user survey may contain both qualitative and quantitative questions.
Qualitative user survey	A **user survey** where questions are answered in free text form. Notes: 1. Qualitative user surveys are used to understand **context of use**. 2. In a qualitative user survey questions are about **users**' experience with the current **interactive system** and their expectations for the new **interactive system**. Examples of questions to understand **context of use**: 1. “What was your business when you last used the car rental website?" 2. "What do you expect from a car rental website?"
Quantitative user survey	A **user survey** where questions are answered by selecting an answer from a choice of alternative replies. Notes: 1. A quantitative user survey can be used to evaluate the **user experience** before, during and after the use of an **interactive system**. Most often, the **user experience** is **evaluated** after the use of the **interactive system**. 2. Often, the questions are statements which the **user survey participant** is asked to rate on a scale like 1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree Can’t answer I don’t want to answer this question (optional) A scale of this type is called a Likert scale. 3. Usually, scales with 3, 5 or 7 steps are used. Avoid scales with an even number of steps because experience shows that **user survey participants** skip questions or abandon the **user survey** if there is no step where they can indicate that they neither agree nor disagree (neutral). 4. The two adjectives at the end points must be definitely opposed. They must be meaningful for each of the aspects that **user survey participants** are asked to rate.

Term	Definition
Questionnaire	A set of questions that is used collect data from **users**, often in a **user survey**. Notes: 1. The process for creating a useful and usable questionnaire is described in the beginning of this section. 2. Questionnaires must be usable. a. The purpose of the questionnaire must be clearly explained at the start. b. A realistic estimate of the time it takes to fill out the questionnaire must be provided at the start. b. Each question must contribute significantly to the purpose of the questionnaire. c. The questionnaire must keep **user survey participants** informed of their progress. d. The first questions in the questionnaire must match the **user survey participants'** understanding of the purpose of the questionnaire. For example, do not start a questionnaire by asking about the **user survey participant**'s age and sex. 3. Questions in questionnaires must be usable. a. Questions must be easy to understand. Speak the **users'** language. Avoid jargon that is unknown to **users**. b. Questions must be unambiguous c. Questions must be free from built-in assumptions. d. Questions should offer **user survey participants** an opportunity to explain their answer in free text form. e. Questions must be positively phrased. Avoid negations. It's difficult for some **user survey participants** to understand the meaning of disagreeing with a negatively phrased question. Example:" I don't think the system is difficult to use." f. Ask one question at a time. Avoid statements like "The trainers did a good job", because it's unclear what to answer if one trainer did a good job and the other did a lousy job.. g. Use the same rating scale throughout the questionnaire. 4. Use standard questionnaires instead of self-invented questions. Examples of standard questionnaires that can be used to measure satisfaction are **SUS** and **UMUX-LITE**. Standard questionnaires can be combined with other questions. Never use only part of the standard questions and leave others out. Standard questionnaires come with specific rules for scoring results, often combined with benchmark data. 5. This definition applies to both digital and paper questionnaires.

4.1. Roles in a user survey

LO #	Learning Objective
4.1.1	Understanding the user survey author role (K2)
4.1.2	Knowing the user survey administrator, user survey evaluator and user survey participant roles (K1)

Term	Definition
User survey author	A **usability professional** who creates, tests and maintains a **questionnaire**.
User survey administrator	A person who administers a **user survey** Note: 1. Administering a **user survey** includes a. **Recruitment** of **user survey participants** b. Sending out **questionnaires** c. Following up on **questionnaires** d. Forwarding **questionnaires** to the **user survey evaluator**.
User survey evaluator	A person who analyzes the responses from a **user survey**.
User survey participant	A person who participates in the **user survey** by filling out the **questionnaire**.

4.2. Examples of standard questionnaires

LO #	Learning Objective
4.2.1	Knowing SUS and UMUX-LITE (K1)

SUS	A simple, ten-item attitude scale giving a global view of subjective assessments of **usability**. Notes: 1. SUS means System Usability Scale 2. The 10 questions in SUS are a. I think that I would like to use this system frequently b. I found the system unnecessarily complex c. I thought the system was easy to use d. I think that I would need the support of a technical person to be able to use this system e. I found the various functions in this system were well integrated f. I thought there was too much inconsistency in this system g. I would imagine that most people would learn to use this system very quickly h. I found the system very cumbersome to use i. I felt very confident using the system j. I needed to learn a lot of things before I could get going with this system 3. Each question is answered on a five step scale with the ending points "Stongly disagree" and "Strongly agree". Only the two ending points are labeled. 4. The SUS has been widely used in the evaluation of a range of **interactive systems**. The scale has been used extensively over a ten-year period and has produced data that allow SUS ratings to be positioned relative to other **interactive systems**. For more information, see http://www.measuringu.com/products/SUSpack
UMUX-LITE	A very simple, two-item attitude scale giving a global view of subjective assessments of **usability**. Notes: 1. UMUX means Usability Metric for User Experience 2. The UMUX-LITE questionnaire has just two questions, which are answered using a seven step scale with the ending points "Stongly disagree" and "Strongly agree". Only the two ending points are labeled. The questions are: a. This system's capabilities meet my requirements. b. This system is easy to use.

5. Model Seminar

This appendix describes a model seminar that teaches the entire curriculum in 3 days.

The stucture of the model seminar is not mandatory; trainers are free to organize their CPUX-UT seminar in any way they consider optimal. The length of the seminar is not mandatory either; trainers may organize seminars lasting 2 or even 1 day, or seminars that last more than 3 days.

Seminar day 1

Start	Topic
09.00	Introduction • Overview of seminar • Presentation of participants and trainer
09.30	Usability evaluation • Overview of usability evaluation methods and when they should be applied • Usability maturity
10.00	Break
10.30	Planning a usability evaluation • Overview of usability test • Usability test plan • Usability test script Recruitment of test participants
11.00	Usability test – Preparation
11.30	>>>
12.00	Lunch
12.30	>>>
13.00	Usability test tasks • Common problems in usability test tasks
13.30	>>>
14.00	>>>
14.30	Exercise 1: Watch videos from usability test, extract findings and build consensus with others on findings
15.00	>>>
15.30	Exercise 2: Evaluate usability test report
16.00	>>>
16.30	Properties of a good usability test report
17.00	End of seminar day 1

Seminar day 2

Start	Topic
09.00	Moderation • Common problems during test sessions and how to avoid them
09.30	>>>
10.00	>>>
10.30	Extract and analyze usability findings • Descriptions and ratings of findings • Recommendations for redesign and improvement
11.00	Communicating findings • The psychology of selling findings to stakeholders • The KJ-method and affinity diagramming • Characteristics of a good usability test report
11.30	Quantitative usability evaluation • Efficiency • User satisfaction
12.00	>>>
12.30	Lunch
13.00	>>>
13.30	Variants of usability test
14.00	Exercise 3: Write usability test script including test tasks
14.30	>>>
15.00	>>>
15.30	Exercise 4: Moderate usability test session
16.00	>>>
16.30	Exercise 5: Recommendations for improving an interactive system
17.00	End of seminar day 2

Seminar day 3

Start	Topic
09.00	User survey • Creation process • Questionnaire • SUS, • UMUX-LITE
09.30	>>>
10.00	Ethics in usability evaluation (5 minutes), followed by Inspection methods
10.30	Inspection methods (continued)
11.00	Exercise 6: Inspection
11.30	>>>
12.00	Lunch
12.30	>>>
13.00	Description of the certification process • Theoretical test • Practical test
13.30	>>>
14.00	Mock exam – theoretical examination
14.30	>>>
15.00	Afternoon break
15.30	Certification, theoretical test (90 minutes)
16.00	>>>
16.30	>>>
17.00	End of seminar

Pre-examination training

To help student candidates pass the practical examination as smoothly as possible, CPUX-UT trainers should offer their students an individual assessment of their skills in writing good usability test tasks.

As part of a CPUX-UT training course trainers should ask each student to submit 4 different tasks for a usability test of an interactive system to the trainer. The trainer must provide individual feedback to each student. In order to use the course time as efficiently as possible, this exercise should be given as a piece of homework to be handed in during or shortly after the course, usually by email.

The interactive system used for this exercise must be a system that would qualify for the practical examination. The Accuweather-example in section 8 of the Examination Regulations (available from www.uxqb.org) may be used.

6. Important changes compared to previous versions

A version of this curriculum with information about all changes made in comparison to the previous version is available from the Editor.

Date, version	Change
04-01-2016, Version 1.03	First version.
12-10-2016, Version 1.06	Minor adjustments and corrections. Added sections: Acknowledgment, Pre-Examination Training (in Appendix 1) Added articles: Positive findings (inspections); Imprecise, lazy (usability test task); Major changes in the following articles: Usability Maturity; Heuristic; Quantitative usability test; Task completion time (previously: Time on task); User survey process Other changes: Some learning objectives changed from “Mastering” to “Understanding”, or from “Understanding” to “Knowing”; Model seminar Minor changes in the following articles: User experience evaluation; Inspection; Heuristic evaluation; Expert usability review; Evaluator (inspection role); Qualitative usability test; Informed consent agreement; Non-disclosure agreement; Usability test session; Preparation of usability test session; Usability lab; Pre-session interview; Moderation; Leading questions; Usability test task scenario; Clues; Analysis of findings; Findings and recommendations (section in a usability test report); Success rate; Failure rate; Disaster rate; Confidence interval; Outlier; Umoderated usability test; Quantitative user survey; Questionnaire; SUS A version that shows changes from version 1.03 in some detail is available from the Editor.
09-02-2017, Version 1.07	Classifications of learning objectives added. Minor changes in the following articles: Usability lab (drawing and associated text); Moderation (order of test tasks); Leading questions (confirmation questions); Usability test task (starting point); Communication of findings (communication must be two-way); Quantitative usability test (satisfaction, assists), Task completion time (uncertainty); SUS & UMUX-LITE (scales) A version that shows changes from version 1.06 in some detail is available from the Editor.

Index