



CPUX-UT Curriculum

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

Version 1.18 EN, 1 November 2020



Publisher: UXQB e.V.

Contact: info@uxqb.org

www.uxqb.org

Table of Contents

Introduction	4
Acknowledgments.....	4
Learning objectives.....	5
1 Overview of usability evaluation	6
2 Usability inspection.....	10
2.1 Approach.....	11
2.2 Usability inspection.....	12
2.3 Heuristic evaluation	13
2.4 Other usability inspection methods.....	16
3 Usability test.....	18
3.1 Prepare usability test.....	21
3.1.1 Usability test plan	21
3.1.2 Usability test script	22
3.1.3 Pilot usability test session.....	22
3.1.4 Recruiting test participants	22
3.2 Conduct usability test sessions.....	26
3.2.1 Preparation of a usability test session	27
3.2.2 Briefing and interview of test participant	30
3.2.3 Moderation	31
3.2.4 Typical problems in moderation.....	32
3.2.5 Usability test tasks.....	34
3.2.6 Typical problems in usability test tasks.....	36
3.2.7 Debriefing.....	39
3.3 Communicate usability findings	41
3.3.1 Basic concepts	41
3.3.2 Involving stakeholders.....	43
3.3.3 Analyse observations and identify usability findings	44
3.3.4 The KJ-method, affinity diagramming	45
3.3.5 Usability test report.....	47
3.3.6 Tips for the usability test report.....	51
3.3.7 Video summary	52
3.4 Roles in a usability test.....	53
3.5 Quantitative usability test.....	56
3.6 Variants of usability test.....	61
3.7 Ethical rules for usability tests	65
3.8 Legal aspects of usability tests.....	67
4 User survey.....	71
4.1 The steps in a user survey	72
4.2 User surveys and questionnaires	73
4.3 Examples of standard questionnaires.....	75
5 Model seminar.....	77

5.1	Seminar day 1	77
5.2	Seminar day 2	78
5.3	Seminar day 3	78
5.4	Pre-examination training.....	79
6	Important changes compared to the previous version	80
7	Index	81

Copyright 2020 The User Experience Qualification Board, www.uxqb.org. The UXQB hereby accords permission to use all or parts of this document for certification and other relevant purposes, provided the source is clearly acknowledged.

Introduction

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

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

The following basic requirements apply for this curriculum:

- It must conform to the established conventions for usability evaluation, as described in popular textbooks and international standards.
- It must conform to the CPUX–F curriculum (available at 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. The description of the model seminar is shown 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 definitions with notes and examples. Each section contains numerous related definitions. The terms shown in boldface in a definition are elaborated in a separate definition. Moreover, the index at the end of the curriculum can be used to quickly locate a definition, which for some terms appear in the CPUX–F curriculum (see Index).

Acknowledgments

The editors express their gratitude towards working group members and reviewers of this and the previous versions of the Curriculum for their constructive contributions in the technical content. Special 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, Amadeus Dilleuth, Anja Endmann, Thomas Geis, John Goodall, Kasper Hornbæk, Rüdiger Heimgärtner, Daniela Keßner, Oliver Kluge, Ludwig Meyer, Knut Polkehn, Matthias Reisemann, Catharina Riedemann, Sabine Rougk, Michaela Thielke, David Travis, Susanne Waßerroth, Melanie Wieland, Chauncey Wilson.

Learning objectives

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

The table at the start of each section presents LOs.

The word “Foundation” in the “LO#” (learning objective number) column indicates that the LO and associated terms are defined in the CPUX–F foundation-level curriculum (available at www.uxqb.org).

LOs are characterised by the following keywords:

Know – recite, recognise

Understand – compare, distinguish, explain, substantiate, summarise

Be able to – analyse, communicate, document, execute, plan

1 Overview of usability evaluation

The purpose of a **usability evaluation** is to identify as many valid **usability problems** and **positive findings** as possible in an **interactive system**.

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.

Usability evaluation can be carried out:

- with **users** – the appropriate methods are **usability test** and **user survey**;
- without **users** – an appropriate method is **usability inspection**.

Learning objectives

Foundation	Understand how and when usability evaluation is used in human-centred design activities
1.1	Understand the differences between usability tests, usability inspections and user surveys
1.2	Understand the selection of the most appropriate usability evaluation method in a context
Foundation	Understand the difference between formative and summative evaluation and how to select the appropriate approach in a context
1.3	Understand quality criteria for a usability evaluation
1.4	Know agile usability evaluation
1.5	Know user experience evaluation and accessibility evaluation
Foundation	Know what usability maturity is

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**).

Selecting a usability evaluation method

The process of determining the most appropriate usability evaluation method in a given context.

The criteria for selecting a usability evaluation method are:

1. The purpose of the usability evaluation.
 - To increase usability awareness or usability maturity in the organisation, run a usability test. This may convince sceptical stakeholders.
 - To evaluate effectiveness, use a qualitative or a quantitative usability test or a usability inspection.

- To evaluate whether efficiency requirements have been fulfilled or whether the usability of a product has measurably improved, use a quantitative usability test.
 - To evaluate whether satisfaction requirements have been fulfilled, use a user survey.
2. Usability maturity.
Use usability test rather than usability inspection if the usability maturity of the organisation is low. Controversial outcomes from a usability inspection can be dismissed as “opinions” in an immature organisation. 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.
Usability inspections are the fastest with respect to test duration; 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.
In early design stages where only low-fidelity prototypes are available for evaluation, use a formative method such as usability inspection, qualitative usability test or user survey.
 5. Resources.
If few resources are available, consider a discount usability test, unmoderated usability test or Rapid Iterative Testing and Evaluation (RITE).
 6. Type of development process – waterfall, agile or other.
See agile usability evaluation.
 7. Mix methods.
 8. It is perfectly OK to mix several recognized usability evaluation methods in one project.
Examples:
 - A quick discount usability test of a new app combined with a few more formal in-depth usability test sessions.
 - A quick usability 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 user experience of the **interactive system**.

User experience and the satisfaction component of usability are evaluated through user surveys.

A user experience evaluation addresses the whole user experience with the interactive system.

Examples:

- Billboards and ads that make users aware of the interactive system.
- Training in the use of the interactive system.
- 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.
- Problems that are not handled by the user interface of the interactive system, for example notifications of delays, handling of complaints, unsolicited calls, etc.

User experience can be evaluated by asking users to keep diaries while they use the interactive system over an extended time period.

Accessibility evaluation

A process through which information about the **accessibility** of an **interactive system** is gathered in order to improve the **accessibility** of the **interactive system**, usually a **usability inspection**.

Accessibility guidelines and requirements can for example be found in

- Web Content Accessibility Guidelines (WCAG) for websites;
- ISO 9241-171 Guidance on software accessibility;
- EN 301 549 Accessibility requirements for ICT products and services;
- the Barrier-free Information Technology Ordinance (BITV; Germany).

Accessibility evaluations are often conducted as usability inspections, which include checking:

- the provision of alternative text for images and the correct mark-up applied to form fields or data tables so that those elements can be interpreted correctly by the user's assistive technology. Semi-automatic evaluation tools are available to do parts of this check;
- the availability of features used by people with disabilities; for example, using only the keyboard, using screen magnifiers, screen readers and other assistive technologies.

Usability tests can be carried out with test participants who use assistive technologies.

Quality of a usability evaluation

The degree to which a **usability evaluation** fulfils commonly accepted requirements for professional **usability evaluation**.

Commonly accepted requirements for professional usability evaluation are:

- described in this CPUX–UT curriculum;
- described in recognised textbooks.

An important step to improving the quality of usability evaluations is to realise that quality could be a problem.

Check for quality problems by:

- applying the CPUX–UT Usability Test Checklist (see “Checklist for the evaluation of the practical test” at www.uxqb.org/en/documents);
- having your approach to usability evaluation reviewed by one or more qualified, neutral outsiders; for example, once every 3 years. Reviews by colleagues are less valuable.

Humility and openness to constructive criticism is useful for increasing quality. Consider reviews and quality assurance opportunities, not nuisances.

Usability testers should pay attention to comments and criticism from peers, that is, other user experience professionals, test participants and stakeholders. It's a common mistake to believe that if your stakeholders criticise 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.

In an agile environment, design teams work in short week- or month-long development cycles, called sprints. In each sprint, the goal is to get a feature or a group of features designed and coded.

Usability evaluation approaches that work well with agile development:

- RITE, Rapid Iterative Testing and Evaluation.
- Weekly testing – test participants are recruited well in advance and scheduled each week; for example, each Tuesday, so that whatever is ready can be usability tested. Appropriate usability test tasks are prepared shortly before the usability test.
- Short discount usability test sessions; for example, in a cafe, conference or trade show, where you have access to lots of potential users. Discount usability test sessions last three to ten minutes.

2 Usability inspection

Usability inspection is a usability evaluation tool that is based on the judgment of one or more evaluators who examine or use an **interactive system** to identify potential **usability problems**. The evaluators also identify deviations from established **dialogue principles, heuristics, user interface guidelines** and **user requirements**.

The evaluators base their evaluation on their experience as **user experience professionals** or as **users** of the **interactive system** that is being evaluated.

Usability inspection techniques are **heuristic evaluation** and **cognitive walkthrough**

Learning objectives

2.0.1	Understand when to use usability inspection and when to avoid it, especially with respect to usability maturity
2.0.2	Understand the steps of a usability inspection
2.0.3	Understand heuristic evaluation
2.0.4	Understand basic heuristics
2.0.5	Know the quality criteria for sets of heuristics
2.0.6	Know the cognitive walkthrough
2.0.7	Understand the differences between usability inspection, heuristic evaluation and cognitive walkthrough

2.1 Approach

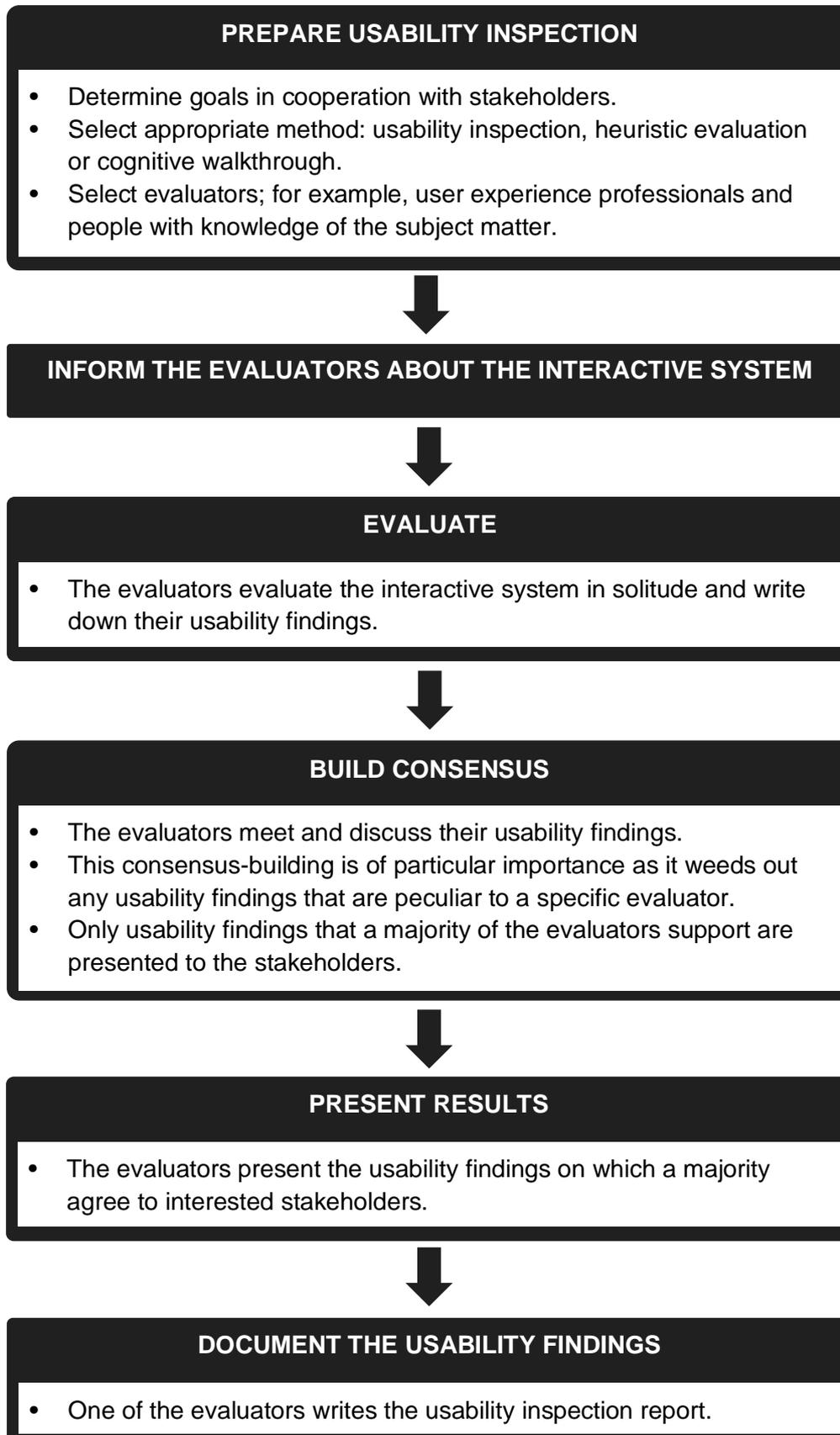


Figure 1. The usability inspection approach

2.2 Usability inspection

Usability inspection

A **usability evaluation** based on the judgment of one or more **evaluators** who examine or use an **interactive system** to identify potential **usability problems** based on **usability inspection criteria**.

Evaluators can be user experience professionals and also people with knowledge of the subject matter, but with little usability knowledge, such as users.

Usability inspection techniques are:

- heuristic evaluation;
- cognitive walkthrough.

Usability inspection is opinion based and may cause opinion wars, which is a serious disagreement where the key arguments are unsubstantiated opinions rather than data or facts. Opinion wars are a sign that an organisation lacks usability maturity or of inexperienced user experience professionals.

The choice between usability test and usability inspection depends on:

- availability of user experience professionals;
- availability of users;
- confidentiality of interactive system to be evaluated;
- usability maturity in the organisation (see also “Selecting a usability evaluation method”);
- available resources: In general, usability inspections are more cost-effective than usability tests.

Frequent mistakes in usability inspections:

- No frank discussion between evaluators of the usability findings. Any usability finding suggested by just one evaluator is accepted uncritically.
For the success of a usability inspection, it is crucial that evaluators are capable of rejecting questionable usability findings and recognising good usability findings, even if they are suggested by one evaluator only.
- Evaluators are not sufficiently familiar with the interactive system and its restraints, which means that usability findings are either not useful for development teams or they are easily brushed aside.
- Evaluators focus on minor details such as user interface guideline violations when there are serious problems with effectiveness and efficiency.

Successful usability inspections:

- discuss the product, not the people responsible;
- focus on causes, not symptoms.

Example:

- Symptom – users can’t find the “Purchase” button because it is in an unusual place.
- Cause – a style guide is either missing or not being observed.

Usability inspection criteria

Generally accepted or agreed rules or rules of thumb for usable **user interfaces**.

Usability inspection criteria include:

- user requirements;
- heuristics and dialogue principles;
- user interface guidelines.

Usability inspection report

A document that describes the **usability findings** from a **usability inspection**.

The structure of a usability inspection report is similar to the structure of a usability test report.

Usability findings from a usability inspection most often are not justified in the usability inspection report. Of course, if the client asks, the evaluators must provide a reasonable justification.

Usability findings from usability inspections are rated as described in the definition of “Classification and rating of usability finding”.

Positive findings are just as important in the results of a usability inspection as they are in the results of a usability test. Please refer to section 3.3.1, “Positive finding”.

2.3 Heuristic evaluation

Heuristic evaluation

A **usability inspection** in which one or more **evaluators** compare an **interactive system** to a list of **heuristics** and identify where the **interactive system** does not follow those **heuristics**.

The list of heuristics must be manageable (see Heuristic). Usually about 10 heuristics are used.

An argument against heuristic evaluation is that it requires evaluators to judge 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.

Frequent mistakes in using heuristic evaluation:

- The evaluation is based on gut feeling rather than heuristics. Usability findings are assigned to one or more heuristics after the usability findings have been found. The correct approach is to let the heuristics drive the heuristic evaluation.
- Evaluators do not fully understand the heuristics. Many heuristics are compact and interpreting them correctly requires some experience.
- Experience shows that many user experience professionals who claim to do heuristic evaluations actually do usability inspection because they report usability findings that could not possibly be found using the heuristics. For example, error messages that are not noticed by users are not covered by Nielsen’s heuristics.

- The evaluators use heuristics made up by themselves, instead of generally recognised heuristics.

Heuristic

A generally recognized rule of thumb that helps to achieve **usability**.

The purpose of a heuristic is to increase the chance that an evaluator will detect usability problems during the usability evaluation of an interactive system.

Quality criteria for reliable sets of heuristics:

- generally recognised – a reliable set of heuristics has stood the test of time;
- comprehensible;
- useful;
- manageable – not too many heuristics; usually about 10 heuristics are used.

The most widely recognised set of heuristics are the 10 heuristics created by Jakob Nielsen. These heuristics are described in the following 10 definitions.

Visibility of system status

The **interactive system** should always keep **users** informed about what is going on, through appropriate feedback within reasonable time.

Examples:

- While moving files, an app regularly updates the message “5955 of 42573 files moved”.
- After a password is changed, the app confirms “Password successfully changed”.

Match between system and the real world

The **interactive 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.

Examples:

- The message “This page does not exist” is easier to understand than “Error 404”.
- A car rental app explains technical terms such as CDW as “CDW – Collision Damage Waiver – exempts the renter from having to pay for any damages to the rental car”.

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.

Examples:

- A user accidentally hits the Exit button. The app offers the choices “Save and exit”, “Exit without saving” and “Cancel”. The availability of the three buttons satisfies this heuristic.
- A file system lets the user undo recent file deletions, moves and renamings.

Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

Examples:

- All cars, irrespective of the brand, have the same ordering of gas, brake and clutch pedal.
- In an organisation, all systems consistently use the same phrases for login, for example, “user name” (not user-id) and “password” (not access key).

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.

Examples:

- Users select dates from a calendar, rather than having to type the dates.
- In a message asking the user to confirm the deletion of a file, the default option is “Do not delete the file”.

Recognition rather than recall

Minimise 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 **interactive system** should be visible or easily retrievable whenever appropriate.

Examples:

- In an address form, a drop-down list displays the names and codes of all 50 US states, so the user does not have to remember the correct spelling.
- A search engine helps users retrace their searches by displaying past searches.

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.

Examples:

- A word-processing system allows inexperienced users to select functions by clicking on menus; experienced users can quickly enter Alt-sequences such as Alt+H+K+C.
- A hotel reservation app allows users to easily change the language and currency.

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.

Examples:

- A bank offers two versions of online banking service: the full version and a more limited and less complex version for people who prefer minimalist design – that is, simplicity.
- The home page of a presidential candidate only shows a large picture of the candidate, the candidate's name, a "Donate" button, a "Sign-up" button and a "Learn More" link.

Help users recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Examples:

- The error message "File names must start with a letter" is more constructive than "Error in file name".
- The error message "The pick-up date (16-Dec-2020) must not be later than the return date (13-Dec-2020)" is more precise than "Something went wrong".

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.

Examples:

- If the user hovers over a technical term such as CDW on a car rental website, a helpful explanation appears; the explanation includes an example.
- A remote control for hearing aids comes with an attractive 2-page getting-started guide.

2.4 Other usability inspection methods

Cognitive walkthrough

A **usability inspection** of a **user interface** in the context of one or more specific **user tasks**.

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:

- Will the user try to achieve the right effect?
- Will the user notice that the correct action is available?
- Will the user associate the correct action with the effect that they are trying to achieve?

- If the correct action is performed, will the user see that progress is being made towards task solution?

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 inspection.

3 Usability test

In a **usability test**, representative **users** perform specific **usability test tasks** one by one with an **interactive system** to enable **user experience professionals** to identify and analyse **usability problems**.

The main activities in a **usability test** are:

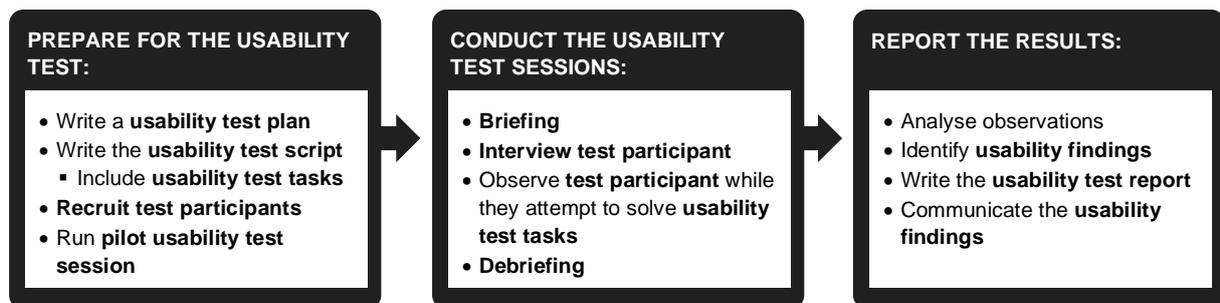


Figure 2. The main activities in a usability test

Learning objectives

- | | |
|-------|---|
| 3.0.1 | Understand the 3 phases and the steps of a usability test. |
| 3.0.2 | Understand the difference in purpose between a qualitative and a quantitative usability test. |
| 3.0.3 | Understand how to conduct usability test for products such as mobile phones, ticket vending machines, TV sets, apps and collaborative applications. |

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**.

A step-by-step overview of a usability test is available in the beginning of this section.

Usability tests are conducted for the following major reasons:

- to assess whether user requirements have been met;
- to uncover usability problems so they can be corrected;
- to convincingly demonstrate to stakeholders that their interactive system contains serious usability problems that can be found and fixed. For more information, see Section 3.3 (the definition of “Selling usability findings”).

Results from a usability test are unique in one aspect: they show what representative users can 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:

- Testing tasks such as “Is the design of the home page appropriate for the CD-Shop?” are opinion based and, thus, inappropriate for a usability test.
- Remarks such as “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?”

The concept of usability test usually refers to a test where the test participant and the moderator are in the same location (“face-to-face usability test”). Other forms are “remote usability test” and “unmoderated usability test”.

A usability test can be a qualitative or a quantitative one.

It 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 completely developed interactive systems.

A usability test is conducted by one or more usability testers. For other roles in a usability test, see Section 3.4.

A qualitative usability test is sometimes referred to as a “think aloud test”.

In a usability test, the moderator works with humans. The moderator must observe the data privacy and ethical rules for working with test participants.

Resource estimates for a usability test:

- A rough resource estimate for a complete moderated qualitative usability test with 5 test participants and 45-minute usability test sessions is 20 to 60 person hours. This test can be done in about 2 weeks, from starting to write the usability test plan to the end of communicating the usability findings.
- Estimates for quantitative usability tests strongly depend on the desired accuracy of results. As a rule of thumb, at least 20 test participants are needed to obtain a reasonable confidence interval. A rough resource estimate for a complete moderated quantitative usability test with 20 test participants and 45-minute usability test sessions is 60 to 120 person hours.

Think aloud

A technique where **test participants** are encouraged to speak out loud what they are thinking while working on **test tasks**.

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.

Therefore, during briefing, the test participant must be encouraged to think aloud by simply saying, “Please think aloud”.

If the test participant does not think aloud, remind them once or twice. If that doesn't help either, 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.

In a quantitative usability test measuring, for example, task completion time, think aloud should be discouraged as it may influence the measurements.

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** behaviour.

Compare Qualitative usability test to Quantitative usability test.

3.1 Prepare usability test

Preparations for the **usability test** start with writing a **usability test plan** describing the goals of the **usability test** and the resources required. After the **usability test plan** has been reviewed and approved, the **moderator** writes the **usability test script**, which includes the **usability test tasks** and checklists for the **usability test sessions**. The **moderator** must be aware of **typical problems in usability test tasks**.

A **pilot usability test session** tests the **usability test script**. A reasonable **number of test participants** for the **usability test** are **recruited** using a **recruiting screener**. Each **recruited test participant** receives a **confirmation**. An **incentive** is often used to compensate **test participants** for their participation in the **usability test**.

Learning objectives

3.1.1	Understand the usability test plan
3.1.2	Understand the factors that influence the number of test participants used in a usability test
3.1.3	Be able to write a usability test script
3.1.4	Understand recruiting, including the recruiting screener and the confirmation process
3.1.5	Understand the factors that influence the choice of the incentive
3.1.6	Know the purpose of a pilot usability test session

3.1.1 Usability test plan

Usability test plan

A brief description of the purpose and extent of a **usability test**.

The usability test plan is intended for the management or a client to decide whether the usability test should be run or not.

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 is acceptable.

The usability test plan includes:

- goals of the usability test;
- identification of the test object: name, version and parts of the object to be evaluated;
- the user group or user groups for the usability test;
- 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);
- number of planned test participants;
- approximate length of each usability test session;
- name of moderator or moderators;
- time plan;
- resource estimate for the usability test;
- how the usability findings will be communicated.

The usability test plan must be short and to the point. Often, one page suffices.

Further details about the usability test such as usability test tasks, method and required software and hardware are provided in the usability test script.

3.1.2 Usability test script

Usability test script

A checklist used by a **moderator** in a **usability test session**.

The checklist contains:

- the preparation for the usability test session – what to do before the test participant arrives;
- briefing instructions;
- questions for the interview of the test participant;
- usability test tasks;
- debriefing questions.

A sample usability test script is included in the sample usability test report.

3.1.3 Pilot usability test session

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.

If serious problems are discovered in the usability test script in a pilot usability test session, correct the usability test script and conduct another pilot usability test session.

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.

Prefer real users as test participants. If they are not available, use colleagues. Avoid using people who designed the interactive system as test participants.

Usability findings from pilot usability test sessions may be included in the usability test report if the usability findings seem valid. For example, if the test participants in the pilot usability test sessions are over-qualified for the usability test and still encounter a serious usability problem in a valid usability test task, this usability finding may be considered valid.

3.1.4 Recruiting test participants

Recruiting

A process for selecting candidates that have the appropriate characteristics to participate in a **usability test**, and scheduling the **usability test sessions**.

A recruiting screener is often used to determine whether candidates have the required qualifications to participate in the usability test.

Test participants from previous usability tests can be re-used as long as you are aware of the bias it may introduce. Consider re-using test participants who are good at thinking aloud and provide frank and apt comments, particularly 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.

Consider using a professional recruiting agency. Check the quality of the agency by asking them to call you in an attempt to recruit you.

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 when you have no relationship to the test participants. To avoid no shows, call up test participants or send them an SMS 24 hours prior to the usability test session, saying, for example, “We look forward to seeing you. Your participation is really important to us”.

For critical usability test sessions, consider recruiting a “floater” – an additional test participant who is used only if the regular test participant does not show up. Note that floaters get paid even if all test participants show up.

Temp agencies can often recruit test participants from their lists of temporary workers, who are often able to participate in usability test sessions during normal work hours. 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 an hour.

Market research companies are great if you want test participants with special qualifications.

During recruiting, explicitly inform test participants if you intend to video or audio record the usability test session.

Recruiting 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**, an **interview** or a **focus group**.

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.

Each additional recruiting criterion narrows the pool of potentially available participants. Prioritise criteria and be prepared to discard less important ones if you can't find sufficient test participants within in a reasonable time.

Exclusion criteria for test participants are user experience or design professionals, media persons (because they might write about the usability test) and people employed by competitors.

Ask all questions that may cause a prospective test participant to be rejected as early in the screening as possible. Have a diplomatic phrase ready to turn them down; 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”.

The user group profile is the basis for the questions in the recruiting screener.

Examples:

- Target group is “frequent flyers”; screening question is “How often do you fly? (reject if <10 flights per year)”.
- Target group is “reasonably experienced smartphone users”; screening question is “Please explain to me what an app is (reject if inadequate explanation)”.
Do not ask “Do you know what an app is?” as many people may over- or underestimate their abilities.

Incentive

A gift or payment given to a **test participant** as a “thank you” or compensation for participating in a **usability test**.

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 say about the interactive system. Traditionally, incentives are handed out after the usability test session has been completed. This could also be wrongly interpreted as the test participant being evaluated – the incentive may be seen as 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 saying, “We are not evaluating you”.

Check if there are price limits for gifts to avoid the appearance of bribery. This is particularly important if test the test participants are publicly employed.

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 the participant’s time less valuable than yours.

If you hand out cash incentives, consider tax regulations.

If a receipt is required, have the appropriate form ready for the test participant to sign.

Typical incentives at the time of writing in Northern Europe are in the range of 30–50€ for a session of about one hour in cases where no special qualifications are required. For test participants with special qualifications, such as lawyers or doctors, it may be up to 10 times higher and test participants may demand that the usability test session be carried out at their place.

Confirmation to test participant

Short, usable information that is sent to the **test participant** ahead of the **usability test session** by email or letter.

The confirmation should include:

- date and time of the usability test session;
- location of the usability test session, including usable information about how to get to the location;
- “We are not evaluating you. We are evaluating the product”;

- name and contact information of contact person, in case the test participant gets delayed or has difficulties finding the location;
- brief explanation on 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”.

Avoid any use of terms that might make test participants uneasy about the evaluation, such as “lab”, “laboratory” or “experiment”. Calm down test participants by informing them that evaluations like these are routine.

Make sure that the information is usable, short and precise. Listen carefully to the feedback from test participants or do a usability test of the confirmation.

Explicitly inform the test participants if you intend to video or audio record the usability test session.

Consider including the non-disclosure and release declaration (NDRD) or the informed consent declaration with the confirmation so that test participants do not have to spend valuable time during the usability test session reading and signing it.

Number of test participants

The number of **test participants** required to obtain reasonably reliable results from a **usability test**.

For a qualitative usability test, 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. Doing so is more efficient than using all your resources to run 15 usability test sessions on the same interactive system version.

Finding all or even 50% of all usability problems in an interactive system is difficult. Unfortunately, it is a common 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; see the discussion in Section 3.5 (the definition of “Confidence interval”).

3.2 Conduct usability test sessions

The **moderator** starts by **preparing the usability test session**, which includes selecting a suitable **test location**, which can be a **usability lab**.

When the **test participant** arrives, the **moderator** briefs the **test participant** about their role in the **usability test**, followed by **interviewing the test participant** to learn more about their background. During **moderation**, the **moderator** hands out written **usability test tasks** to the **test participant** and observes their actions. The **moderator** must avoid **typical problems in moderation**. After the **moderation**, the **moderator** quickly **debriefs** the **test participant**.

Learning objectives

3.2.1	Understand the flow of a usability test session
3.2.2	Understand the criteria for the selection of a suitable test location
3.2.3	Know the usability lab
3.2.4	Understand the preparation of a usability test session
3.2.5	Be able to conduct a briefing
3.2.6	Be able to interview a test participant
3.2.7	Be able to moderate a usability test session
3.2.8	Understand typical problems in moderation
3.2.9	Be able to write usability test tasks
3.2.10	Understand typical problems in usability test tasks
3.2.11	Be able to conduct a debriefing

Usability test session

A part of the **usability test** where one **test participant** carries out representative **usability test tasks** using the **interactive system** or a **prototype** of the **interactive system**.

A usability test consists of many usability test sessions.

Usually, usability test sessions are attended by a moderator and observers, who are often stakeholders, and also a note-taker who records important usability findings. The moderator is the only person who can talk to the test participant during the usability test session. For additional roles in a usability test, see Section 3.4.

In a usability test session, the moderator typically:

- prepares the usability test session, as described in Section 3.2.1;
- greets test participants;
- conducts the briefing;
- interviews the test participant;
- hands out usability test tasks to the test participant;
- observes the test participant during usability test task solution;
- conducts the debriefing.

The length of a usability test session can be anywhere from 5 to 90 minutes. Longer sessions may fatigue or stress out both the test participant and moderator. The length of a moderated usability test session is generally 45–60 minutes and that of an unmoderated usability test session is 15–20 minutes.

Moderators should not moderate for more than eight hours per day, including breaks of at least 15 minutes between usability test sessions.

3.2.1 Preparation of a usability test session

Preparation of usability test session

The activities that happen before the **test participant** arrives for a **usability test session**.

Checklist for the preparation:

- Set the screen resolution and zoom level to agreed conditions.
- Clean up the desktop to ensure no unnecessary applications or document icons are present.
- Lock the task bar: it should be visible so that the time appears on video.
- Turn off unwanted applications such as mail, softphone and automatic backup service.
- 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.
- Arrange the cards with the usability test tasks in the right order; each task should be printed on a separate card.
- Clear the browser cache.
- Prepare the audio or video recording (see below), but do not start recording.

Start preparations early. A useful guideline is “It’s better that you wait 5 minutes for the test participant than the test participant waiting 2 minutes for you to finish the preparations”.

Consider audio or video recording. A simple webcam or screen-recording tool can also be useful. Trial versions of usable software screen recorders are available for free such as by TechSmith. Some video conferencing tools also offer recording features; you can set up a session with yourself and invite colleagues. 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.

You need the informed consent of the test participant to record the usability test session. When using cloud-based software for recording, you may have to sign a data processing agreement with the provider.

Test location

The place where the **usability test** is conducted.

Rules of thumb:

- Observers and test participants should be in separate rooms. If this is not possible, observe the restrictions given in “Separate rooms” below.
- The test location should be undisturbed, unless it is part of the usability test that the test participant is deliberately disturbed.

- Observers and note-takers should be able to observe what the test participant does without disturbing them either through a one-way mirror in a usability lab or via video in a separate observation room.
- The moderator can sit with the test participant (recommended) or in the observation room. If the former is followed, the moderator should sit beside and slightly behind them. The moderator must be out of sight when the test participant looks at the screen, but eye contact must be possible.

Examples of test locations are:

- usability lab;
- two office rooms connected by a video link;
- an office room;
- a room at the test participant's residence or place of work;
- a public place; for example, a cafe (this option is mostly used for quick discount usability test sessions, lasting 10 minutes or less).

It is an important part of a usability test that observers discuss, laugh and cry while observing a usability test session. Also, observers should be able to come and go as they please. Regulations limiting the behaviour of observers in an observation room are a sign that the usability test setup is unsuitable as it does not consider the needs of the primary users. The primary users of any usability test session are the observers.

It is technically possible to let observers observe test sessions from their workplace via screen sharing. A major disadvantage of this approach is that it prevents stakeholders from interacting and discussing while they observe a usability test session.

If stakeholders are invited to observe usability test sessions, choose a test location that is as convenient for stakeholders as possible, rather than a remote usability lab.

Separate rooms: Observers and note-takers should be in one room and test participants in another room. If this is not possible, for example, because the test is in a public place or only one room is available, test participants may become aware of the observers.

In this case, limit the number of observers to one or two and instruct them to:

- keep completely quiet during the usability test session, unless addressed directly by the moderator;
- sit in a place where they are out of the test participant's sight, such as behind the test participant;
- arrive before the test participant arrives;
- stay in the room until the test participant has left.

It is important that observers are able to hear clearly what the test participant says. This applies particularly to 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.

When testing desktop systems, the moderator should have a mouse of their own connected to the test computer, so they can quickly close unwanted popups or operate the recording software.

Usability lab

Two or more rooms that are specially equipped for **usability tests**.

A usability lab often consists of:

- a test room where the test participant sits in front of the display unit. Usually, the moderator sits close but out of the view of the test participant, as illustrated by in Figure 3;
- an observation room where observers (stakeholders) and note-takers can watch test participants from as they solve usability test tasks.

Usually, the two rooms are separated by a one-way mirror that enables observers to watch the test participant but not vice versa. The camera in the test room (marked by K in Figure 3) is connected to a monitor (M) in the observation room. It enables observers to see the test participant's face. This can be done via a webcam located on the test participant's monitor.

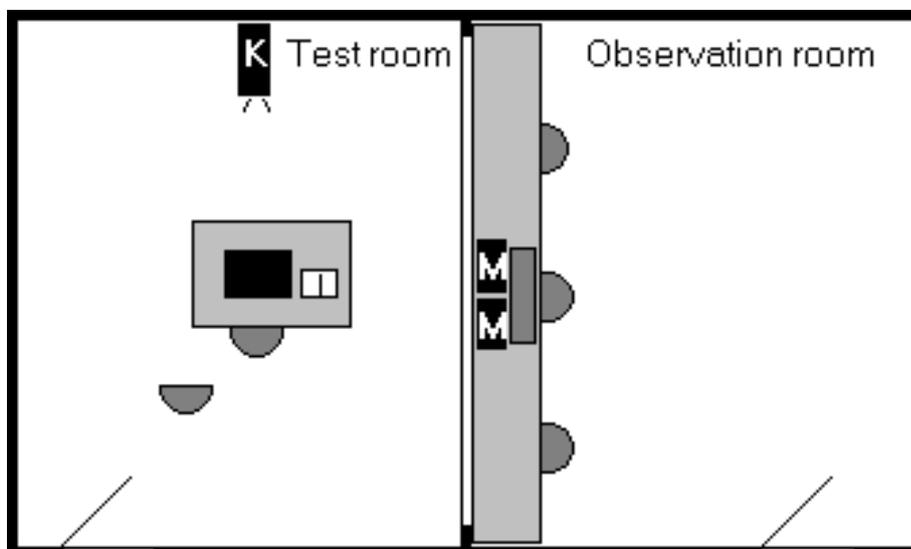


Figure 3. A usability lab

The advantages of using a usability lab are as follows:

- Observers can observe usability test sessions together.
- Usability test sessions are easy to observe for note-takers and observers.
- Usability test sessions are conducted under similar conditions.
- Usability test sessions are easy to video record.
- Observers can enter and leave during usability test sessions.
- It's a great showpiece for a management that wants to demonstrate physically to visitors that the organisation is deeply committed to usability.
- A usability lab enables testing interactions that are not easily digitally transmissible, such as physical interaction with hardware or gesture interfaces.

The disadvantages of using a usability lab are as follows:

- The context is artificial.
- It is expensive to set up and maintain.

3.2.2 Briefing and interview of test participant

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 will be.

Checklist for the briefing:

- Turn off your mobile phone and ask everyone in the room to do the same.
- Briefly introduce yourself and any other people who are in the room.
- Ask test participants to read and sign the Non-Disclosure and Release Declaration (NDRD), if applicable. Tip: send the NDRD and the Data Privacy Statement (DPS) to test participants ahead of the usability test session so that they have time to consider it and so you don't lose valuable time waiting for them to read it during the usability test session.
- Start the recording when you have the permission.
- Hand out the DPS to the test participant and ask them to give consent orally to the DPS and the video recording, so you have the informed consent on video. If you don't use video recording, ask the test participant to sign an informed consent declaration.
- Explain the purpose of the usability test. For example, say "We're asking people to try using this website so we can see if it works as intended".
- Say "We are not evaluating you. We are evaluating the product".
- Say "You are free to leave any time. Even if you leave early you will still get your gift, and we will delete the recording".
- Ask the test participant to think aloud by saying, "Please think aloud so I can understand what you are considering".
- Avoid phrases like "Please tell us what you like and dislike" or "We want your feedback" as they encourage participants to review the product rather than use it as they normally would.
- Finish the briefing by asking "You are welcome to ask questions at any time. Do you have any questions right now?"

A briefing must be efficient and 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.

The moderator should use a written checklist during briefing. It is OK to read the briefing from a carefully worded script, provided that the moderator has frequent eye contact with the test participant.

Interview test participant

An activity in a **usability test session** where the **test participant** answers questions about their background and previous experience with the **interactive system** and related **interactive systems**.

Interview questions should briefly address the following information about the test participant:

- Personal background.
Generally, the profession is of some interest, while age is rarely relevant.
- 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” and “What would you do to get an app on your smartphone?”;

- Previous experience with the interactive system.
Examples: “Are you familiar with this system?” If the answer is yes, ask, “What have you used it for?”, “How was your experience with it?”, “Please tell or show in some detail what you did”.
- Previous experience with similar interactive systems.
Examples: “Are you familiar with similar systems?” If the answer is yes, ask, “Which one?”, “What have you used it for?”, “What was your experience with it?”, “Please tell me what you did”.

The interview should be brief. Do not ask for information that you already have from a reliable recruiter. Focus on information that will make actual difference to your analysis.

Novel or innovative findings from the interview and the debriefing can be communicated as “good ideas” or in a separate section in the usability test report.

If no interview is conducted during the usability test session, provide a summary of test participants’ backgrounds to the observers ahead of the usability test sessions.

3.2.3 Moderation

Moderation

The activity carried out by a **moderator** in a **usability test**.

During moderation, the moderator must:

- hand out each task in writing to the test participant, without reading the task aloud; in some cases, where the spelling of a word in the task may provide a clue to the solution, read the task aloud;
- check the task wording to be concise and clear to minimise reading and understanding effort;
- ask the test participant to silently read the test task and describe in their own words what they are being asked to do; with this, the moderator can check whether or not the test participant fully understands the test task; correct the test participant’s understanding politely if necessary (keep in mind that perhaps your formulation was unclear);
- ask the test participant to start working on the test task;
- observe the test participant while they attempt the task;
- take as many notes as possible keeping in mind that moderation is of top priority; watching the video afterwards is not cost efficient;
- guide test participants when they get completely stuck, usually by moving on to the next test task; helping test participants to complete tasks that cause serious difficulties often provides inadvertent clues; such help should be avoided unless solving the current test task is necessary to solve the following one; if the moderator provides help, it must be limited to absolute minimum;
- guide test participants if they deviate from the test task or the interactive system; for example, to another website that is of no relevance to the current usability test.

The moderator should not disclose the overall number of test tasks; this might upset test participants who only finish a subset.

The moderator should 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 completely clear.

The moderator should speak as little as possible during moderation. Neutral utterings that indicate that the moderator is paying attention such as “OK” or “Aha” are acceptable. The moderator should not say, for example, “You solved the task quickly”, as it could give the wrong impression that we are testing the test participant.

Any kind of instructions must be avoided in connection with usability test tasks as they may provide inadvertent clues.

Start immediately with Task 1. Do not ask the test participant to start by exploring the interactive system for a few minutes and saying what they think. This is not goal oriented and encourages opinions.

The moderator should be the only person talking 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.

“Facilitation” is sometimes used as a synonym for moderation.

3.2.4 Typical problems in moderation

Talkative moderator

A **moderator** who talks too much during **moderation**.

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 clue

An unintended hint provided by the **moderator** that helps a **test participant** to solve a **usability test task**.

While the test participant carries out usability test tasks, the moderator should observe their spoken language or body language to not provide unintended hints about how the test task’s solution or the progress of the test participant.

Examples of clues:

- “Don’t worry about this”, signalling “even though you may think so, this is not important for this task”.
- A faint smile or the moderator’s tone of voice may signal “You have almost solved the task. The answer is on the screen”.

- 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.
- The task description mentions a function or navigation target name that is also used in the user interface.
- 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 debriefing.

Leading question

A question that signals a preference for certain possibilities or attempt to direct the reply in a certain direction.

The moderator must not ask leading questions.

Avoid "confirmation questions" where the moderator asks the test participant to confirm an answer. Confirmation questions considerably reduce the opportunity for surprises (see the fourth example below).

Examples of leading questions:

- "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?"
- "Would you have preferred to have the option of only looking for rental cars with automatic transmission?"
- Any question that starts with "Would you expect ...", for example, "Would you expect the message to appear on top?"
- "Do you always do it like this?"

Bias

Inclination or prejudice for or against a person or system in a way considered to be unfair.

The moderator must be unbiased and neutral. The moderator must not defend the interactive system. The moderator must not express their own views on the interactive system. Further, 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.

Empathy

The ability to understand and share the feelings of the **test participant**.

Test participants tend to blame themselves when they cannot solve a task. To proactively counter this tendency, empathise by saying, "Mind that we're not testing you but the software. Thank you very much for showing us something we can improve!"

3.2.5 Usability test tasks

Usability test task

A description of a **task** that a **moderator** asks a **test participant** to carry out during a **usability test**.

A good usability test task:

- matches the goals of the usability test as defined in the usability test plan;
- is relevant from the test participant's point of view; avoid system-oriented usability test tasks;
- is relevant from the stakeholders' point of view.

For each usability test task, describe:

- the precise phrasing of the task or usability test task scenario handed to the test participant;
- 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;
- why the task was deemed important to the evaluation, including what the task is supposed to evaluate;
- any data given to the test participant for solving the task; for example, a delivery address or information in the database when the test participant starts the task;
- 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".

Each test task must have new diagnostic value. Explain the diagnostic value of each task – what the task is supposed to evaluate. Note that test tasks must differ significantly from each other.

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 successful completion of the previous one.

Prepare enough usability test tasks so that you are sure that you won't run out of tasks within the time allocated for the usability test session.

The first usability test task should be simple but non-trivial so test participants experience a quick success. This is particularly important if a test participant appears a bit stressed by the situation.

The sample usability test report includes a sample set of usability test tasks.

Open-ended usability test task

A **usability test task** that leaves it up to the **test participant** to define parts of the **goal** for the task.

Open-ended tasks are useful because they motivate test participants.

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.

The usability test report must describe the usability test task that each test participant actually decides to carry out in order to help the reader understand what the test participant did.

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.

The usability findings from open-ended usability test tasks are more difficult to analyse than the usability findings from closed tasks because open-ended tasks vary within a given frame.

Examples:

- “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”.
- “Buy whatever you want at this online music store for up to \$50. We will reimburse you for all your expenses”.

Closed usability test task

A **usability test task** that defines the **goal** of the **task** in detail.

A closed task is the opposite of an open-ended task. Both open-ended and closed tasks are useful in usability tests.

Examples:

- “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”.
- “Buy Tchaikovsky’s Nutcracker Suite with Sergiu Celibidache published by Emi Classics”.

Key usability test task

A usability test task whose easy completion is essential to the success of the product.

Test key tasks before testing specialised tasks.

Example:

- On an email website such as Outlook, usability test key tasks such as “send email”, “send email with attachment”, “move email to folder”, etc. before you usability test specialised tasks such as “create signature”.

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**.

The difference between a usability test task and a usability test task scenario is that the latter is contextualised – it contains additional contextual information that justifies the task.

Sample usability test task for a calendar programme:

- “Use the calendar programme 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 programme to notify you a week in advance so you’ll remember to buy a present”.

Scenarios should be minimalistic; don’t invent motivations the test participant would be unlikely to have (see “Pretender task”).

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**.

To create a good set of usability test tasks, proceed as follows:

- At first, don’t look at the interactive system when generating ideas for tasks as this may lead to a bias and cause you to create only tasks that can be solved by the interactive system.
- Ask current and prospective users, the development team, friends and colleagues to suggest appropriate key tasks.
- Get buy-in from the product team by working with team members to create the set of tasks.
- 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.

To improve a set of usability test tasks, proceed as follows:

- Solve your own usability test tasks using the interactive system; discard tasks that turn out to be unreasonable; inform the stakeholders if one or more key tasks turn out to be unsolvable. Unsolvable key tasks should not be included in the usability test.
- Run a pilot usability test session and solicit feedback on the task set from the test participant.
- During the debriefing after the first 2–3 regular usability test sessions, solicit feedback on the task set by asking the test participant “Are these tasks realistic?”
- If needed, modify the usability test tasks; avoid sweeping changes, make the smallest possible changes instead which will solve the problems.

3.2.6 Typical problems in usability test tasks

Clue

Information in a **usability test task** that is unknown to most **test participants** and inadvertently helps them to solve the task.

Examples:

- “Use Outlook to create a personal signature. Having done so, send a short greeting to me (the moderator) using your personal signature”.

Note: “Signature” is a clue here as it is a term used by Outlook. It may not be known or understood by the test participant. The task tests the test participants’ ability to recognise a keyword rather than 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”.

- “Find a song by Liza Minnelli”.

Note: This task contains significant clues, if given in writing. One of its purposes is to check the error tolerance of the search engine, since experience shows that few people are able to spell Liza Minnelli’s name correctly. Instead, phrase the task as “Find a song by an artist whose name the moderator will tell you”.

- “Use the help system to find information about CDW insurance”.

Note: “Help system” and “CDW” are clues.

Both examples in the definition of “Closed usability test task” contain clues: “Car class intermediate”, “CDW” and the spelling of the complex names.

Pretender task

A **usability test task** that asks the **test participant** to pretend to be someone they are not. Doing so creates an invalid context for the task solution and the **test participant** may even perceive it as an insult.

Pretence within reasonable limits is OK for example, “You just arrived in Frankfurt Airport. A friend, who lives in Hamburg, calls you to say that she had to change her plans and that she is leaving tomorrow, so you must see her tonight. Use your smartphone to find the fastest train connection to Hamburg”.

Pretender tasks become problematic when the role-play goes into the political–ideological realm. The participants must then not only play but act against their world views and values (see Examples 3 and 4 below).

Examples:

- “You are a great admirer of Michael Jackson. What’s the name of his latest album?”
Note: This is both a pretender task and an unrealistic one. A great admirer of Michael Jackson would hardly have to look up the name of his latest album.
- “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”.
- “You are enthusiastic about nature and actively involved in environmental protection. Find an article in the newspaper about environmental damages caused by tourists in Nepal”.
- “You are planning a vacation in Portugal. Rent a powerful sports car to get around”.

Silly task

A **usability test task** that attempts to entertain the **test participant**, often by using “humorous” names of persons, places or products.

Silly tasks should be avoided because the test participants’ sense of humour could differ from the moderator’s. Also, silly tasks might hurt test participants’ trust in the moderator and cause them to be less frank in their answers.

Whenever your usability test task description distracts a test participant from the real task, you have failed – even if some test participants appreciate the humour.

Example:

- “Report noise damage caused by Mr. Makin A. Racket, 14 Pandemonium Avenue, 9999 Uproarton” (test of an insurance company’s website).

Derogative task

A **usability test task** that asks the **test participant** to pretend that he or she sometimes behaves foolishly.

Example:

- “Let’s assume that by mistake you have put two identical CDs into your shopping bag. Please get rid of one of them”.

System-oriented task

A **usability test task** that is relevant from the system’s but not the **test participants’** point of view.

System-oriented tasks are “necessary evils” from users’ point of view. They are often subtasks of a relevant task. Therefore, use a relevant usability test task instead which includes the system-oriented task.

Examples:

- “Register yourself on this website” – Use a usability test task instead which requires registration.
- “Use the help-system to find out what the date format is in this system” – 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.

Sensitive personal information

A **usability test task** that encourages or forces the **test participant** to reveal sensitive personal information.

For usability test tasks wherein sensitive personal information is required, provide play data. For example, if the user is required to enter the name, address, etc. for an order, prepare artificial, neutral data such as John Doe, who lives on 25, Main Street, Testtown.

Do not encourage or force test participants to use real data. If they enter real, personal data anyway, stop them immediately. Otherwise the video will contain personal data.

Examples of sensitive personal information are as follows:

- Private address, e-mail address and telephone number.
- Names, addresses, e-mail addresses and telephone numbers of friends or relatives.
- Social security number.
- Bank account number or credit card number.
- The contents of a test participant’s private email inbox.

Imprecise, hazy task

A **usability test task** whose **goal** is unclear, making it difficult to determine when the **task** is completed.

Example:

- “Find and read the cancellation policy” is imprecise, while a similar task “How much does it cost to cancel a reservation?” is sufficiently precise.

3.2.7 Debriefing**Debriefing**

An activity in a **usability test session** where the **test participant** answers questions about their **user experience** and the general impression of the **interactive system’s usability**.

The purpose of the debriefing is to let test participants share their experience in their own words without much prompting.

The debriefing takes place after the test participant has carried out as many usability test tasks as time allows.

Recommended debriefing questions:

- “Which 2–3 things did you like most about the system?”
- “Which 2–3 things are most in need of improvement?”

Also, ask follow-up questions about anything that was unclear to the test participant or the moderator during the session. These would be opinion-related questions. They are helpful in prioritising the usability findings from the usability test.

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?”

Make the debriefing as brief as possible. The standard questions should not take more than 2–3 minutes. Usually, the test participant will have little to add to what they already said during the usability test session.

Refrain from demonstrating the correct solution to a task that caused difficulties, even if the test participant asks for it. Say, “We learned from you that we need to improve this design. It wouldn’t make sense to show you something we need to change anyway”. The test participant may perceive a demonstration as a critique of their performance.

If stakeholders are observing the usability test in real time, invite one or two of 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:

- demonstrating all the wonderful things that the interactive system can do;
- directly or indirectly blame the test participant.

Stop video and audio recording when the debriefing is completed. The scope of the test participants' informed consent is the usability test session, not any arbitrary time period afterwards.

3.3 Communicate usability findings

During this activity, the **moderator**, **note-taker** and **communicator** analyse usability findings, which is either a **usability problem** or a **positive finding**.

The **usability findings** are discussed with interested **stakeholders** using the **KJ-method** or **affinity diagramming**. **Involving the stakeholders** in this activity helps **selling the usability findings**.

The results are documented in a **usability test report**, which consists of an **executive summary** and lists of **usability findings**, which are **classified and rated**. Optionally, a **recommendation for improvement** is provided.

A **video summary** may be used to communicate some of the most important findings to interested **stakeholders**.

Learning objectives	
3.3.1	Understand the causes for communication problems and how to prevent them
3.3.2	Understand the importance of involving stakeholders in the planning and execution of the usability test
3.3.3	Understand the techniques for selling findings to stakeholders; for example, the KJ-method and the underlying principles
3.3.4	Be able to analyse usability findings
3.3.5	Know ways of communicating usability findings
3.3.6	Be able to write a usability test report with substantial usability findings
3.3.7	Understand classification and rating of usability findings
3.3.8	Know the video summary

3.3.1 Basic concepts

Usability finding

A result from a **usability evaluation**.

A usability finding can describe:

- a usability problem;
- something that users liked – a positive finding;
- a good idea from a test participant;
- a new user requirement resulting from the analysis of usability findings.

Notably, findings must be anonymous; it must not be possible to associate a test participant with a usability finding.

A classification and rating should also be assigned to each usability finding.

Usability findings from a usability test should be based solely on what test participants are able to accomplish with the interactive system when carrying out representative usability test tasks. Usability findings should not include opinions, neither from the moderator nor the test participant. Personal opinions regarding potential usability problems are valid results from usability inspections.

It's OK to report usability 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 findings must be substantial. Avoid trivial usability findings such as "Too many ads" or "Test participants liked the home page". Turn such findings into non-trivial findings by explaining how ads interfered with test participants' use of the website or how the home page helped them specifically to get their job done.

Usability problem

A difficulty in using the **user interface** that affects the ability of the **user** to achieve their **goals effectively, efficiently** or with **satisfaction**.

Usability problems can lead to confusion, error, delay or outright failure to complete some **task** on the part of the **user**.

Positive usability finding

Something that **test participants** liked about the **user experience** or that clearly helped them achieve their **goals**.

As a rule of thumb, at least 25% of the reported usability findings should be positive.

If at first the moderator can only think of few things that the 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.

Reporting positive findings:

- ensures that the features liked by the test participants are not removed simply because the development team was not aware of the appreciation;
- creates a more positive attitude towards the usability study.

Positive findings must be completely positive. If a positive finding is not completely positive, separate it into a positive finding and a usability problem.

3.3.2 Involving stakeholders

Communication of usability findings

The process of informing **stakeholders** and others about the **usability findings** from a **usability test**.

The following techniques are available for this process:

1. Informal discussions of usability findings with stakeholders; for example, between usability test sessions.
2. Workshops with stakeholders.
3. Presentations of usability findings with subsequent discussion.
4. Writing and distributing a usability test report with subsequent discussion.
5. Creating and distributing a video summary.
6. Entering the usability findings in the bug database for the interactive system.
Communicators should understand the rules that are in effect for entering bugs and follow them carefully.

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 usability findings before they are distributed to others and correct any misunderstandings. After distributing the usability findings, discuss them with stakeholders.

In order to decide on appropriate communication techniques, the usability maturity of the organisation should be considered:

- If usability maturity is low, use appropriate techniques for selling usability findings.
- If usability maturity is high, use the most efficient techniques (points 4 and 6 in the above list).

Regardless of the technique used, listen carefully to feedback from the stakeholders. Avoid one-way communication or superficial dismissal of objections to the usability findings.

Selling usability findings

The process of convincing **stakeholders** in organisations with low **usability maturity** that **usability findings** from a **usability test** should be taken seriously and acted upon.

Some stakeholders are sceptical towards usability and usability tests as expressed by the anonymous quote “Cowboy programmers don’t need no stinkin’ usability”.

Some designers view their user interface as an extension of themselves and thus take it personally when someone finds fault with it.

Avoid opinions as there is no magic answer to the question “Why are your opinions better than mine?” and it can easily lead to an “opinion war”. Opinion wars only have losers, therefore, be careful with opinion-based methods such as usability inspection and heuristic evaluation. Stakeholders, in particular product managers and designers, are skilled in discussing opinions. User experience 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.

Involve stakeholders in the planning and execution of the usability test (see the definition of “Involving stakeholders”).

Involving stakeholders

A part of the process of making acceptance of **usability problems** smoother.

The most important stakeholders are the people who decide the changes from the usability test which should actually be implemented and the people who do the implementation.

Some ways of involving stakeholders and making them feel that the usability test is also “their baby” are by:

- inviting stakeholders to participate in writing the usability test plan and the usability test script, particularly the usability test tasks;
- inviting them to participate in the recruiting process, particularly defining the test participant profile and creating the recruiting screener;
- inviting and encouraging them to observe usability test sessions;
- involving them in determining the usability findings through use of the KJ-method.

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:

- 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.
- Conduct usability test sessions at locations that are convenient for stakeholders, for example where many stakeholders work, instead of a faraway usability lab.

3.3.3 Analyse observations and identify usability findings

Analysis of usability findings

The process that extracts **usability findings** from observations during **usability test sessions**.

Proceed as follows:

- During the usability test session, the note-taker must record usability observations, usually by writing them down, reflecting events that cause problems with or have a positive effect on effectiveness, efficiency and satisfaction.
- After each usability test session, the note-taker and the moderator should meet to discuss the observations from the usability test session while everyone still remembers what happened.
- After all usability test sessions have been completed, the moderator and the note-taker should separately extract 20–30 usability findings and 5–10 positive findings each from their observations. These usability findings reflect the observations that they consider most important.

- The moderator, the note-taker and the communicator meet and have a frank discussion about their usability findings. The usability 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, especially if there are many different usability 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 usability findings must be limited. For example, if you find 75 usability problems, you can't 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 prioritise the usability findings and report only the ones that are most important from a usability point of view.

The discussions between the moderator, the note-taker and the communicator must be 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”).

Compare the usability findings from the current usability test to any relevant previous usability test reports.

Usability findings should be based on similar observations of at least two test participants.

- 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.
- If just one test participant encounters a serious or critical problem and a few other test participants have visited the page or window where the problem occurs, the moderator should consider potential harm and use their experience, experience from others and heuristics to judge whether the problem should be reported or not.

3.3.4 The KJ-method, affinity diagramming

KJ-method

A brainstorm-based method for quickly establishing consensus among **stakeholders** regarding the most important **usability findings** from a **usability test**.

The KJ-method consists of the following steps:

1. Encourage all stakeholders to observe one or more usability test sessions and take notes.
2. Invite all stakeholders who have observed at least one usability test session. Conduct the KJ-session immediately after the final usability test session.
3. 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.
4. Display the cards; for example, on a wall or a blackboard. Self-adhesive post-it notes are useful for this purpose.
5. 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.

6. 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.
7. Name each group. Use cards of a different colours for group names. Use the group names as section headings in the usability test report.
8. 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/

The KJ-method is essentially a brainstorm, particularly the first part (KJ-session Step 3–5). In order not to inhibit the brainstorm, discussions are not allowed during this part.

The main strengths of the KJ-method are:

- The usability findings are defined by the people who will implement them – the stakeholders. This increases the acceptance of the usability findings.
- The usability 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.
- All stakeholders get heard.

Positive findings are not considered in a KJ-session.

The optimal number of active participants in a KJ-session is 3 to 8.

The communicator organises 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.

Stakeholders who have not observed at least one usability test session must remain passive because their contributions are opinions.

This method is named after the Japanese ethnologist Jiro Kawakita.

Affinity diagramming

A hierarchical technique for organising 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.

When used to analyse usability findings, affinity diagramming is similar to the **KJ-method**.

3.3.5 Usability test report

Usability test report

A document that describes the results of a **usability test**.

The report must be effective and efficient for the key stakeholders, particularly the development team and managers who make decisions about what will be changed.

A usability test report must be efficient. Specifically, this means that it must be:

- short and succinct – report at most 50 usability findings regardless of the size of the usability test; the recommended number of usability findings is 5–25 (see the tip Usability test report size);
- uniform – all usability test reports from an organisation should have the same structure and graphic layout;
- comprehensible – avoid usability jargon like “mental model” and “WCAG” (Web Content Accessibility Guidelines).

A usability test report should contain the following sections:

1. Executive summary.
2. Table of contents.
3. Usability findings and recommendations.
4. Identification of the test object: name, version and parts of the object that were evaluated (if applicable).
5. Purpose of the evaluation, including listings of or references to relevant user requirements.
6. Evaluation method.
7. Target group for the interactive system.

Section 1, 2 and 3 must appear first and in the given order, while the remaining sections may be placed in one or more appendices.

A usability test report is always required for a usability test, even if you use the KJ-method, even if you have little time and if you use agile usability evaluation. However, the rules above are quite flexible. A simple conforming report might be made in Microsoft PowerPoint format with one slide on executive summary, two to three describing the five most important usability findings, and three in total on the object of evaluation, the purpose of the evaluation, the usability test tasks and the evaluation method.

The following information should normally not be included in a usability test report:

- An explanation of what usability is.
- A description of the human-centred design process.
- Details about recruiting; for example, copies of the recruiting screener and the confirmation sent to test participants.
- Transcripts – extensive and verbatim text on what one or more test participants said during usability test sessions.

ISO standard 25066, “Common Industry Format (CIF) for Usability – Evaluation Report”, contains recommendations for usability test reports.

A free sample usability test report that illustrates the requirements in this curriculum is available at www.uxqb.org.

Executive summary

A section in a **usability test report** that provides a brief overview of what was **usability tested** and the most important **usability findings** from the **usability test**.

The target audience for the executive summary is managers who need a brief overview of the most important usability findings from the usability test.

The executive summary should not exceed one page and should contain brief descriptions of:

- the object of evaluation;
- when the test was conducted;
- the purpose of the evaluation;
- the evaluation method;
- 2–4 most important positive findings;
- 2–4 most important usability problems;
- general recommendations based on the usability findings (optional).

Examples of general recommendations:

- Review of the conformance of error messages to a set of agreed guidelines;
- Review of all UI texts for clarity and brevity.

Usability findings and recommendations (section in usability test report)

A section in a **usability test report** that describes the 5–50 most important **usability findings** from the **usability test** and associated **recommendations for improvement** of the **interactive system**.

The description of each usability finding should include:

- classification and rating of usability finding;
- a header that briefly describes the usability finding;
- an indication of the number of test participants who encountered a problem, for example “a few”, “most” or “all” test participants;
- a description of the usability finding;
- relevant quotes from test participants during their encounter of the usability finding (optional);
- recommendations for improvement (optional);
- screenshots illustrating the usability finding (optional).

Evaluation method (section in usability test report)

A section in a **usability test report** that describes how the **usability test** was conducted.

Enough information should be provided to determine the appropriateness of the method and assess the validity of the results as well as enable replication.

This section should include:

- 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;
- information regarding the physical and technical environment in which the usability test took place;
- usability test script;
- an anonymised list of test participants;
- name and email addresses of the usability testers who conducted the usability test;
- for a quantitative usability test, the data analytic treatment used.

Classification and rating of usability finding

A measure given to a **usability problem** from a **usability test** to indicate the type of the **usability finding**, its impact, criticality and consequences on the **user experience**.

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.

Possible classifications are:

- usability problem – Each usability problem must have a severity rating as described in the following note;
- positive finding – Something that worked well in the context of the current usability test or that test participants liked;
- good idea – A suggestion from a test participant that could lead to a significant improvement of the user experience;
- functional problem – Bug.

Typical severity ratings of usability problems are:

- minor problems – Minor dissatisfactions, noticeable delays or superficial difficulties.
Example: Test participants used 20 to 30 seconds to locate the Search bar;
- major problems – Substantial delays or moderate dissatisfaction
Example: Test participants were unable to locate Search because it is unobtrusive; they eventually found the information by other means;
- critical problems – Test participants gave up, showstopper, substantial dissatisfaction or minor financial damage to user;
Example: Test participants were unable to solve key tasks because they were unable to locate Search;
- catastrophic problems – Existential threats, potentially life-threatening, bodily harm or substantial financial damage to the user or to the organisation;
Example: Test participants did not notice that the liability insurance for the rented car provided insufficient coverage; they could be ruined if they caused a serious accident.

Usability problems should only be classified as catastrophic after consultation with the risk manager or a top-level manager, otherwise the user experience professional oversteps their authority.

Important parameters that influence severity ratings are:

- frequency – How often does the usability problem occur?

- impact – How badly does it hit the user and the user’s environment when it occurs?
- persistence – How quickly will users learn to avoid the usability problem?

Avoid severity ratings like “Imperative to fix as soon as possible” or “Must fix, regardless”. Usability testers are responsible solely for usability, not for the cost to fix. Comments of this type overstep the charter of the user experience professional.

Recommendation for improvement

An optional suggestion for how a **usability problem** can be solved.

Arguments for including recommendations in a usability test report:

- Recommendations help stakeholders understand what the usability problem really is. In other words, the recommendation extends the problem description.
- Many stakeholders expect such advice from user experience professionals: “That’s why we hired you!”
- Missing recommendations create a vacuum, which is often filled with unusable solutions.

Arguments against including recommendations:

- The solution is obvious or the communicator knows the stakeholders well enough to trust that they don’t really need help, or worse, might actually find it insulting.
- Unfamiliarity with all the constraints (technical, business, legal, political, etc.).
- Lack of confidence that the recommendation is a good one.
- The problem is large enough in scope that the communicator can’t easily describe a solution in a few sentences.

Example: Major problems in understanding the navigation structure.

Complex problems are best solved by team collaboration, which also increases buy-in; that is, the chance that the recommendations will actually be implemented.

Emphasise that recommendations are just one of the several ways to solve a usability problem. Start your recommendations by stating, for example, “One way of solving this problem is to ...”

Means for expressing recommendations are:

- text;
- altered screenshots, possibly with callouts.

A great recommendation:

- solves the problem;
- is technically feasible;
- recommends the least possible change which leads to the intended outcome;
- is illustrated, for example, with an altered screenshot;
- is justified;
- provides detail;
- speaks the readers’ language;
- is constructive and direct;
- is short;
- addresses only the original usability problem.

3.3.6 Tips for the usability test report

1. **Size:** The quality of a usability test report is not proportional to the number of reported usability findings. On the contrary, it is important that communicators prioritise usability findings and only report the essential usability findings, independent of the extent of the usability test.
2. **Sample report:** A CPUX–UT Usability Test Report Example produced by the UXQB is available free of cost at www.UXQB.org, in English and German. 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.
3. **List of test participants:** Provide a table with participants (row) and key characteristics (columns) of test participants which are relevant to the validity of the usability test. This could include:
 - profession or job title;
 - previous knowledge of and interest in the interactive system being tested;
 - previous knowledge of and interest in similar interactive systems;
 - age or age group;
 - sex (note that sex is often of less importance).

Do not provide names of test participants – not even given names – or any other information that could identify them.

Provide meaningful information about profession or job title. For example, “manager”, “student” or “retired” are not helpful, while “HR manager in a medical company”, “graduate student, organic chemistry” and “butcher, retired” are helpful.

Describe any differences in key characteristics between the test participant sample and the actual user population; for example, “Actual users might attend a training course whereas test participants were untrained”.

4. **Include examples in general usability findings.**

Example:

 - Usability findings such as “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”.
5. **Tactful usability findings:** Usability findings should be expressed tactfully. Rude usability findings create enemies, not more usable interactive systems.

Examples of rude usability findings:

 - “The availability of the books is not shown!!!”
 - “The design of the home page is unprofessional and sloppy”.
6. **One issue per usability finding:** Usability findings must not be conglomerates of several independent but possibly related usability problems.
7. **Use icons to show severity ratings in usability test reports.** Use icons that are intuitive and considerably contrasting so that readers can easily locate them; for example, all usability

problems that have the ratings of critical or catastrophic (see the sample usability test report for more examples).

3.3.7 Video summary

Video summary

A video that illustrates some of the most important **usability findings** from a **usability test** through appropriate video clips from **usability test sessions**.

The purpose of the video summary is to briefly present the most important usability 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.

The length of a video summary should be 3–15 minutes.

Include at least one positive finding in the video summary.

Illustrate each usability problem or positive finding with at least 2 clips where different test participants experience the usability finding to show that more than one test participant encountered the usability problem.

Start the video by displaying a title for long enough (suggestion: 8 seconds) so that viewers have ample time to read the essential information, which is:

- object of evaluation;
- test date;
- name of the communicator who created the video summary;
- where to find the usability test report; provide, for example, an intranet link.

Use subtitles or title slides at the start of the clips for a usability problem or positive finding to briefly explain the point.

Avoid “talking heads” – videos of the communicator explaining what the usability finding is.

Appropriate video clips can also be integrated into the usability test report.

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 result in usability improvements.

3.4 Roles in a usability test

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

Learning objectives

3.4.1	Understanding the moderator, note-taker and communicator roles
3.4.2	Understand the observer, administrator and test participant roles

Role

A duty or function that a person carries out within an organisation.

A role describes a set of connected behaviours, rights, obligations and norms in a job situation.

Roles are occupied by individuals, who are called “actors”.

Depending on the complexity of a project, several people might share one role or multiple roles might be assigned to one person.

Work roles are “hats” that people wear when they take on the corresponding job responsibilities and perform associated activities.

Example:

- In a usability test session, one person can simultaneously occupy the roles of moderator and note-taker. Contrastingly, in another usability test session, one person may act as moderator and two other people may both act as note-takers.

Usability tester

A neutral **user experience professional** who evaluates **user interfaces** in various stages of realisation.

Usability tester is a generic **role** covering the **roles moderator, note-taker and communicator**.

Moderator

A neutral **user experience professional** who conducts a **usability test session**.

The moderator’s responsibilities during a usability test session are described under usability test session.

Not everyone has the empathy and patience to be a good moderator.

The moderator should remain curious and never leave unsure about what happened.

Examples of well-placed curiosity:

- During the interview in a test of a video streaming platform the test participant says, “I just cancelled my streaming subscription”. The moderator asks out of curiosity, “Why did you do that?”
- During the test of a train ticket website, the test participant mentions that he prefers using the site on a mobile device. The moderator asks for an example of when he last did that to learn about the context of use.
- During task solution, the test participant does not click a button that would have solved the task. During the debriefing, the moderator asks whether the test participant noticed the button and if so, how the test participant interpreted the button.

Every moderator should try at least once to be a test participant. Further, to ensure an unbiased usability test, the moderator must be neutral. For example, product owners and developers are unsuitable as moderators.

The moderator does not have to be technically proficient. Neutrality, knowledge of how to deal correctly with the test participants and familiarity with the procedure for a usability test are more important than the technical knowledge of the test object. The moderator must conduct pilot usability test sessions until they have the necessary knowledge of the test object.

The term facilitator is sometimes used as a synonym for the term moderator.

Note-taker

A **user experience professional** who makes notes of **usability findings** during a **usability test session**.

Having a note-taker allows the moderator to fully concentrate on the test participants.

Note-taking is often handled by the moderator to keep the costs down.

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, while others prefer to speed-type on a laptop. However, the noise of the keyboard can be annoying and may constantly remind the test participants that their actions are observed and recorded. Therefore, hand-written notes are preferred.

The note-taker also plays an active role in the analysis of the usability findings.

Observer

A person who watches **test participants** carrying out the **usability test tasks** during a **usability test session**.

Observers are often stakeholders.

Observers do not interfere in the moderation.

Observers may be actively involved in the analysis of the usability findings, for example, when the KJ-method is used.

Communicator

A **user experience professional** who communicates the **usability findings** from a **usability test**, for example by moderating a **KJ-session** or by writing and presenting the **usability test report**.

The communicator also plays an active role during the analysis of the usability findings.

Administrator

A person who administers the **usability tests**.

Administrative tasks include scheduling the test sessions, test participant management, communication with the 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

A **quantitative usability test** focuses on obtaining the figures for the **effectiveness**, **efficiency** or **satisfaction** of an **interactive system** by measuring **task completion time** (a measure of efficiency), **success rate** (a measure of **effectiveness**), **failure rate** (a measure of **effectiveness**), **disaster rate** (a measure of **effectiveness**) and **satisfaction** which is measured with the help of a **user survey**.

There is a risk of making **invalid measurements**. Further, measurements are subject to uncertainty which is expressed in terms of **confidence intervals**. **Confidence intervals** can be used to compare the performance of **interactive systems**.

Learning objectives

3.5.1	Understand the quantitative usability test
3.5.2	Understand the parameters that are most often measured in a quantitative usability test
3.5.3	Understand confidence intervals and how to use them to compare the performance of interactive systems
3.5.4	Understand the number of test participants in a quantitative usability test
3.5.5	Understand failure rate and disaster rate
3.5.6	Understand invalid measurements and how to handle them

Quantitative usability test

A **usability evaluation** that focuses on obtaining figures for the **effectiveness**, **efficiency** or **satisfaction** of an **interactive system**.

Quantitative usability tests are used to do the following:

- Evaluate a design against quantitative user requirements. The evaluation process is described in the CPUX-F curriculum.
- Compare interactive systems, for example, competitive products.
- Compare versions of an interactive system.

The parameters that are most often measured in quantitative usability tests are the following:

- Task completion time – that is, the time to complete a usability test task (a measure of efficiency).
- Success rate (a measure of effectiveness).
- Failure rate (a measure of effectiveness).
Note: Failures include disasters.
- Disaster rate (a measure of effectiveness).
- Satisfaction.

Success rate + Failure rate = 100%

To plan and analyse the measurements from a quantitative usability test, some knowledge of the following basic statistical concepts is required:

- Mean.
- Geometric mean.
- Standard deviation.
- Confidence interval

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.

When reporting quantitative results it is important to include the confidence interval. There is a tremendous difference between the measurements 100, 110, 120 seconds and 10, 110, 210 seconds even though they have the same mean.

To obtain valid measurements, it is of particular importance to recruit representative test participants.

An attempt to solve a usability test task during a quantitative usability test can have the following outcomes: Success, failure, disaster or invalid measurement.

Participants should be briefed that time is being measured and that they should report when they feel they have solved a task or give up. Add “This is not to put pressure on you – we just want to know how long things take”.

Assists like “Please read the task again” are permissible. More extensive assists like “The answer is not on this screen” are not permissible.

The qualitative usability findings discovered as part of a quantitative usability test should also be reported to help stakeholders determine how their unsatisfactory performance can be improved.

Thinking aloud should be discouraged during a quantitative usability test as it may influence measurements.

Satisfaction is measured with the help of a user survey, for example, SUS.

Task completion time

End time minus start time for a **test participant's** completion of a **usability test task**.

Start time: The start time is not the time when the test participant receives the usability test task. The timing should start after the user has read the instructions and at the point when information on the interface is presented to the user.

The 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 is not sufficient that the answer appears on the screen.

The reported task completion time is the mean value of all the task completion times for all successful solutions of the usability test task.

The mean task completion times from quantitative usability tests with different success rates are not easily comparable.

Determining the exact start time and end time can be difficult. The uncertainty is often 5 or more seconds.

Examples of uncertainty are the following:

- The test participant starts working on the usability test task, but then returns to the instructions and reads them again carefully.
- 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 are successful. The success rate is a measure of the **effectiveness** of the **interactive system**.

Success: A task attempt is successful if the test participant manages to find an acceptable answer without any assistance from the moderator.

It is important to 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.

A task attempt is a failure if the test participant does the following:

- Gives up.
- Is unable to find an acceptable answer within reasonable time
Note: A common time limit is 10 minutes.
- Finds an incorrect answer and believes it to be correct. Compare this with disaster rate.

Failures include disasters.

It is important to 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.

An attempt to solve a usability test task is disastrous if

- the test participant finds a seriously wrong answer and considers it correct – that is, they would continue their work using the wrong answer;
- the consequences of using the wrong answer are critical or an existential threat.

The important distinction between failure (as defined in failure rate) and disaster needs to be noted. Usually, the number of disasters is much smaller than the number of failures.

Examples:

- Disaster for a council website: the test participant finds a wrong date for the next collection of bulky waste.
- Failure for a council website: the test participant finds an outdated telephone number for city services. This is not a disaster since users can easily find out that the number

is wrong when they call it and so the consequences of this from the user's point of view would be limited.

Invalid measurement

A **usability test task** measurement that cannot validly be included in the results of a **quantitative usability test** because of such reasons as equipment malfunction, software malfunction, procedural error, **test participant** error or irrelevant interruption.

Examples of reasons for declaring the measurement of a usability test task invalid:

- The moderator helps the test participant in completing the 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.
- The moderator intervenes to solve a technical problem.
- Hardware or software problems are encountered while using the computer.
- The required basic software is not installed on the computer. For example, the appropriate Adobe Reader is missing.
- The test participant is interrupted for more than 10 seconds (unless the interruption is part of the usability test task). For example, there is an interruption by a colleague or a telephone call.
- The test participant solves or starts solving a wrong usability test task.
- The test participant misunderstands the usability test task or does not read it with sufficient care.
- The test participant starts discussing with the moderator which invalidates the measurement if the discussion lasts more than 10 seconds.

Confidence interval

An estimate of a range of values that includes the true value.

The purpose of the confidence interval is to show with some certainty the interval where the correct value lies, 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.

All the 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 was between 90 and 210 seconds with 90% certainty.

Increasing the number of representative test participants taking a quantitative usability test decreases the length of the confidence interval.

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.

The computation methods for the confidence interval are different for success/failure/disaster rates, task completion times and satisfaction scores. For additional information including regarding calculators for computing CIs and more, visit Jeff Sauro's website www.measuringusability.com

Confidence intervals can be used to compare the performance of key tasks on two interactive systems, A and B.

- Measure the completion times for a key task on A and B.
- Compute the confidence intervals for the key task on A and B.
- Non-overlapping confidence intervals as shown in Figure 4, indicate that the performance of the key task on A is better than the task's performance on B within the chosen uncertainty.

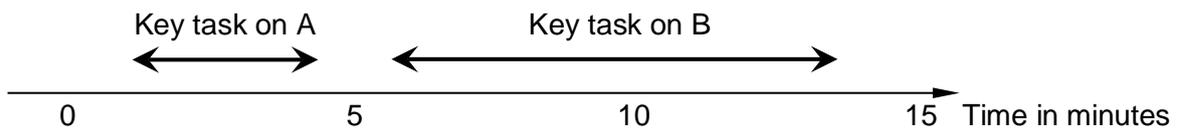


Figure 4 Non-overlapping confidence intervals

- If the confidence intervals overlap as shown in Figure 5, the apparent difference may be mere chance. The reason is that the true value for the key task on A might be at the extreme right of the interval, while the true value for B might be at the extreme left of the interval for B. Thus, the true value for the task on B might be less than the true value for the task on A even though it seems counterintuitive at first glance.

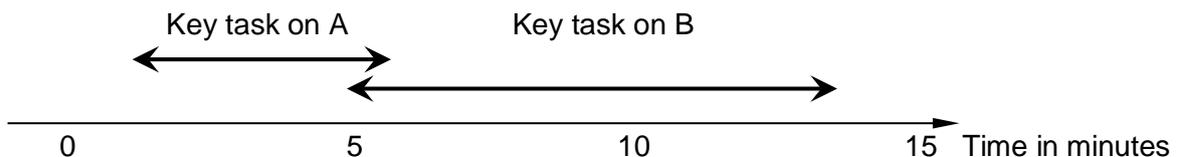


Figure 5 Overlapping confidence intervals

3.6 Variants of usability test

There are several common variants of the face-to-face **think aloud usability test** which are the following: **discount usability test, co-discovery, remote usability test, unmoderated usability test, retrospective recall, Rapid Iterative Testing and Evaluation (RITE), eye tracking, international usability test** and **A/B testing**.

Learning objectives

3.6.1	Understand the discount usability test and how it differs from an ordinary qualitative usability test
3.6.2	Know the co-discovery, remote usability test, unmoderated usability test, A/B testing and RITE
3.6.3	Know the important differences between remote usability test, unmoderated usability test, A/B testing, RITE and ordinary qualitative usability test
3.6.4	Know retrospective recall
3.6.5	Know eye tracking
3.6.6	Know international testing

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**.

A discount usability test has one or more of the following characteristics:

- The moderator, note-taker, communicator and administrator are the same person.
- No usability lab is used; usability test sessions take place, for example, in a meeting room.
- The usability test is limited to 5 test participants.
- The usability test report is short and lists a limited number of usability findings (say, 15). The goal is to write an effective and efficient usability test report.
- The usability test takes less than 30 person-hours to complete which includes the communication of the usability findings.

A discount usability test may include a usability inspection.

Co-discovery

A **usability test session** where two **test participants** carry out **usability test tasks** together.

Co-discovery makes thinking aloud unnecessary because talking to each other comes naturally for the test participants in this case.

Co-discovery requires twice as many test participants as the ordinary usability tests, and some people find working with strangers annoying especially if their problem-solving techniques differ. So despite its advantages, co-discovery is rarely used in practice.

Further, multiple test participants may create privacy issues. Hence, the test participants must sign an NDRD which covers information they learn from the other participant.

Remote usability test

A **usability test** where the **test participant** and the **moderator** are at different locations.

The moderator observes the test participant using an internet connection.

The moderator communicates with the test participant over the telephone or using an internet connection.

This should be compared to usability test and unmoderated usability test.

Unmoderated usability test

A **usability test** where **test participants** solve the **usability test tasks** without a **moderator**.

The main advantage of an unmoderated usability test is that the usability test sessions can be run in parallel and independently of a moderator's schedule, location and time zone. However, the analysis effort is the same as in a traditional usability test.

Test participants' actions are usually video recorded to be analysed later.

Unmoderated usability tests are often conducted on the test participant's computer at the participant's home. Video recording of the usability test is done with the help of the recording software installed on the computer by a vendor of the unmoderated usability test services. After each usability test session, the video recording is automatically transmitted to the vendor and the client for analysis.

Unmoderated usability tests are offered by several vendors who recruit the test participants and conduct usability test sessions using your usability test script for a reasonable fee.

Important quality parameters for 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.

"Unmoderated usability test" is sometimes referred to as "crowd usability test" even though it does 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**.

Watching and commenting on the usability test session video takes approximately as much time as running a new usability test session with a fresh test participant. Before employing this method, carefully consider if the retrospective recall will provide sufficient insight to justify the time investment.

Retrospective recall may be required when you don't want the test participant to think aloud for one or more of the following reasons:

- You are doing a quantitative usability test.
- The test participant is involved in tasks where distraction is potentially life-threatening; for example, air traffic control, factory assembly lines or medical systems.
- Tasks require so much concentration that test participants will be overtaxed if asked to think aloud while working.

Retrospective recall is not an observation of actual test participant behaviour. It may be influenced by irrelevant factors, such as post-rationalisation and attempts to justify non-optimal behaviour.

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.

The RITE method focuses on an instant redesign to fix problems and the confirmation that the solution works with new test participants.

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 stakeholders decide if they will make any changes before the next test participant is tested. The changed user interface is then tested with the remaining test participants.

Eye tracking

Tracking and recording **test participants'** eye movement during a **usability test**.

The purpose of eye tracking is to get an understanding of where test participants look on the screen and where they do not look.

Eye tracking often generates a heat map showing where test participants looked over a period; for example, while solving part of a usability test task.

Eye tracking does not show the test participants' thought activity.

Eye tracking rarely generates important insights that could not have been achieved through an ordinary usability test. However, the usability 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.

The purpose of an international usability test is to understand users' reactions to an interactive system outside the country where it was built and its usability tested.

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 participants. 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.

A and B can be competing designs and each version is presented 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.

The purpose of A/B testing is to validate that a new design or change made to an element in an interactive system is improving the interactive system before the production version is modified.

A/B testing offers concrete numbers based on numerous users. The caveats for uncertainty in quantitative usability tests, in particular, confidence intervals, apply. In contrast, comparative usability tests give qualitative insights into usability differences between two interactive systems based on a limited number of users.

A/B testing is also known as split testing.

3.7 Ethical rules for usability tests

Learning objectives

- | | |
|-------|--|
| 3.7.1 | Understand why ethical rules for usability tests are necessary |
| 3.7.2 | Understand the ethical rules and tips for usability tests |

Ethical rule

A statement that describes a commonly accepted norm for decent behaviour

Ethical rules are necessary to ensure that user experience professionals do no harm to other human beings and that they respect privacy, confidentiality and anonymity.

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>

If you are a member of other professional associations such as the American Psychological Association, additional or different ethical rules may apply.

The ethical rules that are most important for usability testing are as follows:

1. Ensure that the test participant feels safe and important.
 - At the end of the usability test session, the test participant must feel at least as comfortable as at the start.
 - Emphasise that we are never evaluating the test participant, but the interactive system.
 - Maintain a professional neutral demeanour. Don't indulge in sarcasm or flirt with the participants.
2. Demystify the usability test by telling the test participant in advance what to expect.
 - Avoid words such as "lab", "test subject" and "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.
 - Never put any direct or indirect pressure on a test participant to reveal personal information such as their name and address and the contents of their email inbox.

Advice on how to apply the ethical rules:

6. Respect

- The moderator must be respectful towards the test participant, stakeholders and the client.

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 expected”.

- The moderator must pay attention to the test participant’s 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.

7. Giving up

- If the test participant gives up on a task, make it clear that it’s not their fault.

Example: Test participant: “I failed this one”. Moderator: “No. There is no failure in this one. Because knowing this difficulty helps us understand what the usability problems are”.

8. 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.

9. 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.

10. Inappropriate jokes

- Jokes such as “If you can’t figure out how to fill out this form, you will not get your gift” are unethical.

3.8 Legal aspects of usability tests

Before a **usability test session** starts, the **test participant** signs a **non-disclosure and release declaration (NDRD)**. The **NDRD** informs **test participants** that they are working with confidential information and allows the organisation to freely use any **intellectual property (IP)**, such as good ideas that may come out of the **usability test session**. Through **informed consent**, the **test participant** acknowledges that data is being collected during the **usability test session**.

During the **usability test session**, the **moderator** is responsible for ensuring the **data privacy** – that is, protecting any **personal data** that **test participants** disclose – following a **data privacy statement (DPS)**, which is handed over to the **test participant** at the beginning of the **test session**.

Disclaimer: The UXQB does not provide any guarantee for the completeness and correctness of the information in this section. For more information, please consult your legal advisor. We recommend the article “General Data Protection Regulation (GDPR) and user research” by Troeth and Kucharczyk, published on Medium. This article is not part of the present curriculum.

Learning objectives

3.8.1	Understand the Non-Disclosure and Release Declaration (NDRD) and informed consent
3.8.2	Know fundamental data privacy for the protection of personal data, the Data Privacy Statement (DPS), and intellectual property protection measures.

Intellectual Property (IP)

The intangible results of creativity, such as patents, copyrights etc.

To protect IP, file any relevant patent claims before usability testing an interactive system with external users, and request test participants to sign a non-disclosure and release declaration.

Examples of IP:

- Innovative concepts that are invented by the interactive system’s author.
- Innovative ideas voiced by test participants.

Non-Disclosure and Release Declaration (NDRD)

A declaration in which a **test participant** declares that they will not disclose any ideas and concepts they learned in a **usability test session** and that they release **intellectual property** pertaining to any ideas they voice in the **usability test session** to the organisation.

Use an NDRD if the information disclosed to the test participant during the usability test session is confidential or if idea generation is a goal of the usability test session; for example, in early prototype testing.

“Release intellectual property” essentially means that test participants allow the moderator’s organisation to use any ideas they voice in the session.

An NDRD is a unilateral declaration by the test participant. It is not an agreement. The common term “Non-Disclosure Agreement (NDA)” is therefore incorrect.

An NDRD is specific to a usability test session and must identify it by topic and date.

Legal disputes with test participants over usability tests are rare. However, if NDRDs are missing or incomplete, this can negatively impact an organisation’s defence.

Data Privacy

The protection of any **personal data** that **test participants** disclose in the context of a **usability test session** – when arranging the session, during the session and afterwards.

The collection and handling of personal data are subject to strict legislation under the GDPR in the European Union (EU) and the California Consumer Privacy Act (CCPA) in the USA.

The common principles of such legislation are:

- *Purpose limitation.* The collection of personal data is only allowed for a defined purpose, which is defined in a Data Privacy Statement (DPS). Re-purposing personal data, for example, for marketing purposes, is not allowed.
- *Minimalism.* You must only collect and process data that are strictly necessary for the stated purpose.
- *Access to collected data* must be restricted to the smallest circle of named individuals.
- *Retention of personal data* is only allowed until the purpose of the data collection is fulfilled. Also, a maximum retention period must be defined. In usability testing, the period is often 90 days.
- *Rights of the test participant.* Test participants are entitled to access, rectify and purge their personal data.

Personal data

Any information about a particular person, or information that can be used to identify the person.

Retain personal data only until the purpose of the data collection is fulfilled.

- Recruiting forms are obsolete when the test session has been completed and test participants have been reimbursed.

- Test session recordings and informed consent declarations related to the test session are obsolete when the usability test is completed, and results reported.
- NDRDs are needed as long as intellectual property needs to be protected.
- Several data storage products afford the definition of access rights and retention policies to facilitate both data sharing and access restriction. They also automate the deletion of personal data. This prevents personal data from being forgotten and illegally stored.

Protect test participants' anonymity.

- If possible, separate recruiting and usability test session processes organisationally and technically to avoid unwanted "contamination" of test records with personal data. For example, combining test session scheduling information with recording timestamps may identify a test participant.
- Avoid collecting any identifying information during the usability test session, that is, there should be no names on record. Have test participants sign any relevant forms outside the test room or electronically on a separate computer. Do not address the test participant by name while audio is being recorded.

Examples of personal data:

- Name, gender and occupation of test participants are personal data if they can be used to identify a particular person. The job title "software engineer" is not considered as personal data in its own right; however, "lead software architect, 50 years old, female" may very well identify a person at an organisation. Therefore, the job title in this context must be considered personal data.
- Test participants' signatures and their names on declaration documents.
- Audio and video recordings from usability test sessions.
- Transcripts and observation notes that identify the test participant.

Data Privacy Statement (DPS)

A statement that discloses the ways an organisation gathers, uses, discloses, and manages a test participant's data

The DPS is handed over to the test participant before or at the beginning of the test session.

The DPS describes:

- who is collecting the data;
- what data is being collected and why;
- how long the data is stored;
- who may access the data;
- the test participant's data protection and information rights;
- how the test participant can exercise those rights; for example, providing the moderator's name and contact information

The test participant does not have to sign the DPS if there is video documentation that the DPS printout has been handed over to the participant. Displaying the DPS on-screen while recording proves which version was used.

Test participants are entitled to a printed copy of the DPS.

Informed consent

Permission granted in full knowledge of the possible consequences.

The moderator can document that they have obtained informed consent to the data collection in one of the following ways:

- Having the test participant sign an Informed Consent Declaration (ICD) during the briefing.
- Recording on video that the test participant has been informed of the data collection and video recording policies and is not obliged to remain in the usability test session. If the participant leaves the usability test session before its conclusion, they revoke their consent, and any personal data collected so far must be deleted immediately.

4 User survey

User surveys are used to evaluate **user satisfaction** with an **interactive system**.

In a **user survey**, **users** submit subjective data through a **questionnaire** based on their experience of using an **interactive system**. The **usability** of a **questionnaire** is significant; for example, the questions in the **questionnaire** must be comprehensible and the **questionnaire** must keep the **users** informed of their progress.

Learning objectives	
4.0.1	Understand the steps in a user survey
4.0.2	Understand the qualitative user survey
4.0.3	Understand the quantitative user survey
4.0.4	Understand the difference between qualitative and quantitative user surveys
4.0.5	Understand the questionnaire
4.0.6	Understand the usability criteria for questionnaires

4.1 The steps in a user survey

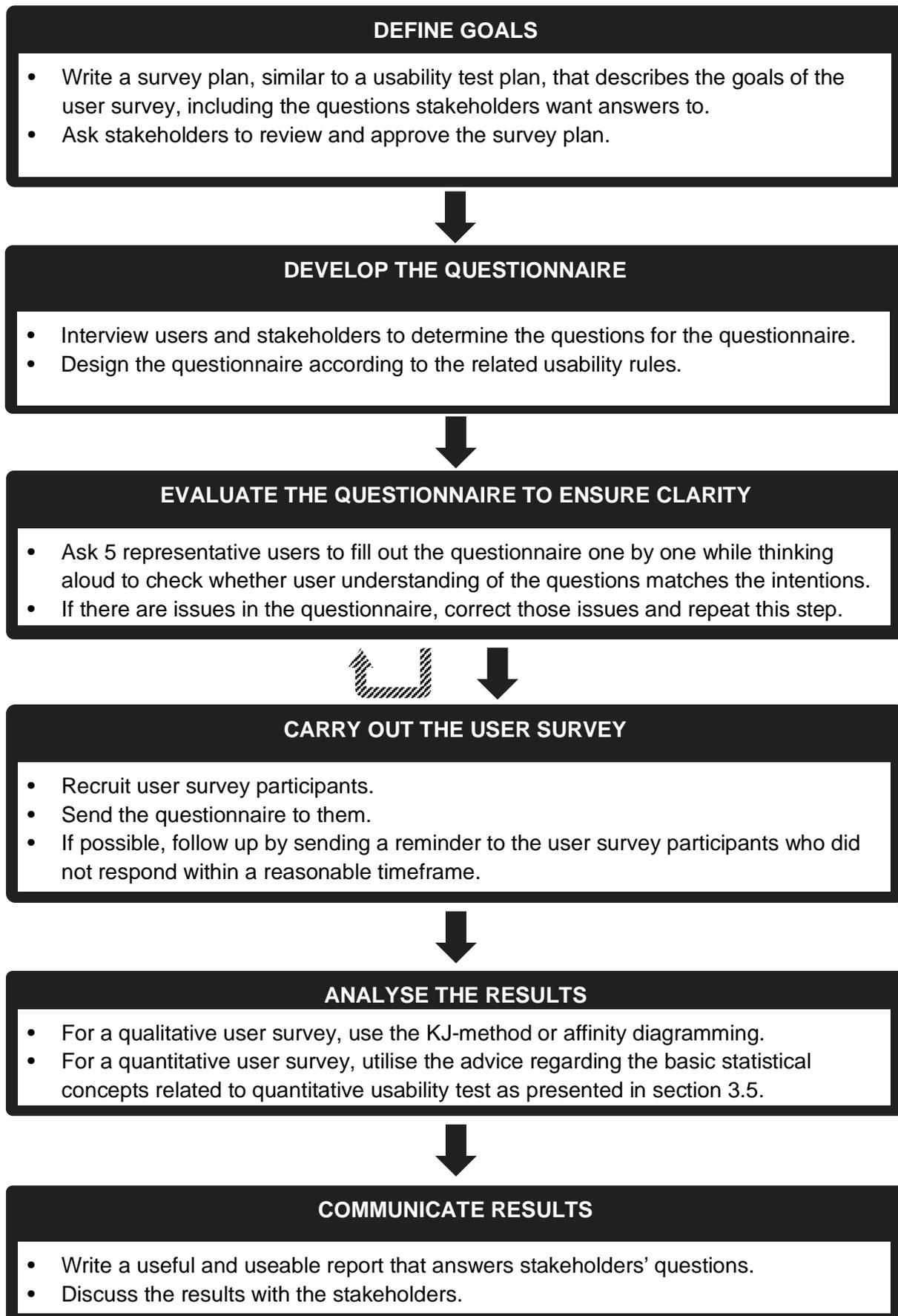


Figure 6 The steps in a user survey

4.2 User surveys and questionnaires

User survey

An **evaluation** where **users** are asked to provide subjective data in a **questionnaire** based on their experience in using an **interactive system**.

User surveys are used to evaluate user satisfaction or user experience with an interactive system and to gather information about the context of use.

Possible user survey goals are as follows:

- Objectifying the levels of satisfaction or dissatisfaction with an interactive system.
- Measuring the effects of changes in an interactive system through user surveys before and after the change.
- Comparing two interactive systems.

The process for executing a useful and usable user survey has been described in the previous section.

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.

A user survey may contain both qualitative and quantitative questions.

Qualitative user survey

A **user survey** through which questions are answered in free text form.

Qualitative user surveys are used to understand the context of use.

In a qualitative user survey, questions focus on the users' experience with the current interactive system and their expectations for the new interactive system.

Examples of questions to be posed to understand the context of use are as follows:

- “What was your business when you last used the car rental website?”
- “What do you expect from a car rental website?”

Quantitative user survey

A **user survey** through which questions are answered by selecting an answer from a set of alternative replies.

A quantitative user survey can be used to evaluate the users' experience before, during and after the use of an interactive system. Most often, user experience is evaluated after the use of the interactive system.

Often, participants are asked to rate statements on a scale such as the following:

- 1 – Strongly disagree
- 2 – Disagree
- 3 – Neither agree nor disagree
- 4 – Agree
- 5 – Strongly agree
- Can't answer

This scale is called a Likert scale.

Do not assign numerical values to each rating and then take an average of the total. There is no rationale behind assigning numbers to responses. The number of responses in each category can simply be added up and presented as a bar graph.

Usually, scales with 3, 5 or 7 responses are used. Avoid scales with an even number of responses because survey participants tend to skip questions or abandon user surveys altogether if they cannot give a neutral response.

The two adjectives at the endpoints must oppose each other. They must be meaningful for each aspect a user survey participant is asked to rate.

A quantitative user survey mainly contains closed questions.

Questionnaire

A set of questions that is used to collect data from **users**, often in a **user survey**.

Questionnaire usability criteria are as follows:

- The purpose of the questionnaire must be clearly explained at the start.
- A realistic estimate of the time it would take to fill out the questionnaire must be provided at the start.
- Each question must contribute significantly to the purpose of the questionnaire.
- The questionnaire must keep the users informed of their progress.
- The initial questions in the questionnaire must match the users' understanding of the purpose of the questionnaire. For example, do not start a questionnaire by asking the user's age and sex.

The usability criteria of the questions in the questionnaire are as follows

- The questions must be comprehensible. The language must be understandable by the users. Avoid jargon that is unfamiliar to the users.
 - For example, research shows that some users had difficulties in understanding the word "cumbersome" in the SUS-question, "I found the system very cumbersome to use."
- The questions must be unambiguous.
 - Example: "Where do you like to shop?" Here, does "shopping" refer to shopping for grocery or gas or online shopping? Does "like" depend on the price, convenience or courteousness of the staff?
- The questions must be free of built-in assumptions.
 - Example: In the question "What do you think of the style of the website?", the scale goes from 1 – Traditional (☹) to 5 – Modern (☺). Here, the smileys are assumptions.

- The questions should offer users an opportunity to explain their answers in a free text form.
- The questions must be positively phrased. Avoid negations. It may be difficult for some user survey participants to understand how to answer a negatively phrased question.
 - Example: In the question “I don’t think the system is difficult to use.”, what does it mean to disagree with this statement?
- Ask one question at a time.
Avoid statements such as “The trainers did a good job” because it’s unclear how a user must answer if one trainer did a good job and another did a lousy job. Instead, ask one question per trainer.
- Use the same rating scale for all the questions in a questionnaire.

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 while leaving others out. Standard questionnaires come with specific rules for scoring results that are often combined with benchmark data.

This definition applies to both digital and paper questionnaires.

4.3 Examples of standard questionnaires

Learning objectives

4.3.1 Know SUS and UMUX-LITE

SUS

A simple, ten-item attitude scale providing a global view of subjective **usability** assessments.

SUS refers to System Usability Scale

The 10 questions in SUS are as follows:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Each question is answered on a five step scale with the ending points “Strongly disagree” and “Strongly agree”. Only the two ending points are labelled.

The SUS has been widely used to evaluate a range of interactive systems. The scale has been extensively used for over ten years and has produced data that allow SUS ratings to be used for other interactive systems.

For more information, visit <http://www.measuringu.com/products/SUSpack>.

If you make changes to SUS, comparisons to the SUS-data from other sources may no longer be valid.

UMUX-LITE

A very simple, two-item attitude scale providing a global view of subjective assessments of **usability**.

UMUX means Usability Metric for User Experience.

The UMUX-LITE questionnaire comprises just two statements that are answered using a seven-step scale with the endpoints of “Strongly disagree” and “Strongly agree”. Only the two endpoints are labelled.

The statements are as follows:

1. This system’s capabilities meet my requirements.
2. This system is easy to use.

5 Model seminar

This chapter is not examinable. It outlines a model seminar that teaches the entire curriculum in just 3 days.

The structure of the model seminar is not fixed; trainers are free to organise their CPUX-UT seminar in any way they consider optimal. The length of the seminar is not fixed either; trainers may organise seminars that may last for two days, one day or even more than three days.

5.1 Seminar day 1

Start	Topic
09:00	Introduction <ul style="list-style-type: none"> • Overview of the seminar • Presentation of the participants and the trainer
09:30	Usability evaluation <ul style="list-style-type: none"> • Overview of the usability evaluation methods and when they should be applied • Usability maturity
10:00	Break
10:30	Planning a usability evaluation <ul style="list-style-type: none"> • Overview of the usability test • Usability test plan • Usability test script • Recruiting test participants
11:00	Usability test – Preparation
11:30	>>>
12:00	Lunch
12:30	>>>
13:00	Usability test tasks <ul style="list-style-type: none"> • Common problems in usability test tasks
13:30	>>>
14:00	>>>
14:30	Exercise 1: Watch videos from a usability test, extract usability findings and build consensus with others on usability findings
15:00	>>>
15:30	Exercise 2: Evaluate a usability test report
16:00	>>>
16:30	Properties of a good usability test report
17:00	End of seminar day 1

5.2 Seminar day 2

Start	Topic
09:00	Moderation <ul style="list-style-type: none"> • Common problems during test sessions and how to avoid them
09:30	>>>
10:00	>>>
10:30	Extract and analyse usability findings <ul style="list-style-type: none"> • Descriptions and ratings of usability findings • Recommendations for redesign and improvement
11:00	Communicating usability findings <ul style="list-style-type: none"> • The psychology of selling usability findings to stakeholders • The KJ-method and affinity diagramming • The characteristics of a good usability test report
11:30	Quantitative usability evaluation <ul style="list-style-type: none"> • Efficiency • User satisfaction
12:00	>>>
12:30	Lunch
13:00	>>>
13:30	Variants of usability test
14:00	Exercise 3: Write a usability test script including test tasks
14:30	>>>
15:00	>>>
15:30	Exercise 4: Moderate a usability test session
16:00	>>>
16:30	Exercise 5: Recommendations for improving an interactive system
17:00	End of seminar day 2

5.3 Seminar day 3

Start	Topic
09:00	User survey <ul style="list-style-type: none"> • Creation process • Questionnaire • SUS • UMUX-LITE
09:30	>>>
10:00	Ethics in usability evaluation (5 minutes) Usability inspection methods

10:30	Usability inspection methods (continued)
11:00	Exercise 6: Usability inspection
11:30	>>>
12:00	Lunch
12:30	>>>
13:00	Description of the certification process <ul style="list-style-type: none"> • Theoretical test • Practical test
13:30	>>>
14:00	Mock exam – Theoretical examination
14:30	>>>
15:00	Afternoon break
15:30	Certification – Theoretical examination (90 minutes)
16:00	>>>
16:30	>>>
17:00	End of seminar

5.4 Pre-examination training

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

As part of the CPUX-UT training course, trainers should ask each student to submit 4 different tasks for the usability test of an interactive system. The trainer must provide individual feedback to each student with regard to their submission. To use the course time as efficiently as possible, this exercise should be given as homework to be submitted 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 the previous version

6.1 Version 1.18, dated 01-11-2020

The previous version was 1.07, dated 09-02-2017.

Learning objective “Mastering” renamed as “Be able to”.

The levels of the following learning objectives were changed from M to U: 2.0.2; 3.0.1; 3.1.1; 3.1.4; 3.2.1; 3.2.5; 3.2.6; 3.2.11; 3.4.1. For LO 3.1.7, U was changed to K.

The following section was added: 3.8. Legal aspects of usability tests

The following definition was added: Accessibility evaluation

The following sections and their associated learning objectives were removed: 2.1 Roles in an inspection; 4.1 Roles in a user survey

The following definitions were removed:

UX Maturity (now covered by CPUX-F); Usability review, Expert usability review; Expert; (Test tasks for) Experienced users; The four types of Desirable usability test tasks; Sunshine task; (Usability test role) Host; Mean; Geometric mean; Standard deviation; Outlier

Major changes were made to the following existing definitions:

- Heuristic: Split the definition of heuristic into 11 definitions with one definition per heuristic; added examples for each of the 10 heuristics
- Recruiting screener: Added examples of screening questions for specific target groups
- Briefing: Clarified when to start the video recording
- Sensitive personal information: Described substituting play data for sensitive personal information
- Classification and rating of usability finding: Added examples of ratings of usability problems
- Confidence interval: Added an explanation of how to use confidence intervals to compare the performance of tasks on two interactive systems
- User survey: Removed step 3.c in the overview of the user survey process

Renamed key terms: “Recruitment” to “recruiting”; “Pre-session interview” to “interview test participant”; “post-session interview” to “debriefing”; “finding” to “usability finding”; “usability professional” to “user experience professional”

Added a short overview at the start of sections 2, 3 and 4. Changed the layout to conform to the new standard for CPUX curricula.

Many minor changes, movements of text and deletions were carried out to accommodate the new definitions and the major changes without increasing the total length of the curriculum. Many tips were integrated into the corresponding definition; some duplicate texts were eliminated.

7 Index

- A/B testing, 64
- Accelerators, 15
- Accessibility evaluation, 8
- Administrator, 55
- Aesthetic and minimalist design (heuristic), 16
- Affinity diagramming, 46
- Agile usability evaluation, 9
- Analysis of usability findings, 44

- Bias, 33
- Briefing, 30

- Catastrophic problem, 49
- Classification
 - Of usability finding, 49
- Closed usability test task, 35
- Clue, 32
 - Usability test task, 36
- Co-discovery, 61
- Cognitive walkthrough, 16
- Communication of usability findings, 43
- Communicator, 55
- Confidence interval, 59
- Confirmation question, 33
- Confirmation to test participant, 25
- Conglomerate, 51
- Consent. *See* Informed consent
- Consistency and standards (heuristic), 15
- Context of use. *See* CPUX-F
- CPUX-UT seminar, 77
- Creation of usability test tasks, 36
- Critical problem, 49
- Crowd usability test, 62

- Data privacy, 68
- Data Privacy Statement (DPS), 69
- Debriefing, 39
- Derogative usability test task, 38
- Dialogue. *See* CPUX-F
- Dialogue principle. *See* CPUX-F
- Disaster rate, 58
- Discount usability test, 61
- Documentation (heuristic), 16
- DPS. *See* Data Privacy Statement

- Effectiveness. *See* CPUX-F
- Efficiency. *See* CPUX-F
- Empathy, 33
- Error prevention (heuristic), 15
- Error tolerance. *See* CPUX-F
- Ethical rules, 65
- Evaluation. *See* Usability evaluation; User experience evaluation ; Accessibility evaluation
- Evaluation method (section in usability test report), 48
- Examples in usability findings, 51
- Executive summary, 48
- Eye tracking, 63

- Facilitation. *See* Moderation
- Facilitator. *See* Moderator
- Failure rate, 58
- Findings. *See* Usability findings
- Flexibility and efficiency of use (heuristic), 15

- Focus group. *See* CPUX-F
- Formative usability evaluation. *See* CPUX-F
- Functional problem, 49

- Gift. *See* Incentive
- Goal. *See* CPUX-F
- Good idea, 49

- Hazy usability test task, 39
- Help and documentation, 16
- Help users recover from errors (heuristic), 16
- Heuristic, 14
- Heuristic evaluation, 13
- Human-centred design. *See* CPUX-F

- Icons for ratings, 51
- Imprecise usability test task, 39
- Inadvertent clue, 32
- Incentive, 24
- Informed consent, 70
- Inspection. *See* Usability inspection
- Instructions. *See* Briefing
- Intangible results, 67
- Intellectual property (IP), 67
- Interactive system. *See* CPUX-F
- Intercultural usability test. *See* International usability test
- International usability test, 63
- Interview. *See* CPUX-F, *See* Interview test participant, Debriefing
- Interview test participant, 30
- Invalid measurement, 59

- Invitation to test participant, 25
- Involving stakeholders, 44
- IP. *See* Intellectual property
- Key usability test task, 35
- KJ-method, 45
- Lab. *See* Usability lab
- Leading question, 33
- Likert scale, 73
- List of test participants, 51
- Location. *See* Test location
- Low-fidelity prototype. *See* CPUX-F
- Major problem, 49
- Match between system and the real world (heuristic), 14
- Measuring usability. *See* Quantitative usability test
- Mental model. *See* CPUX-F
- Minimalist design, 16
- Minor problem, 49
- Model seminar, CPUX-UT, 77
- Moderation, 31
- Moderator, 53
 - Bias, 33
 - Clue, 32
 - Empathy, 33
 - Leading questions, 33
 - Talkative, 32
- NDRD. *See* Non-Disclosure and Release Declaration
- Non-Disclosure and Release Declaration (NDRD, 68
- Note-taker, 54
- Number of test participants, 25
- Observation room, 29
- Observer, 54
- Open-ended usability test task, 34
- Participant. *See* Test participant
- Persona. *See* CPUX-F
- Persona based review, 16
- Personal data, 68
- Personal information, 38
- Pilot usability test session, 22
- Positive finding, 49
- Positive usability finding, 42
- Post-session interview. *See* Debriefing
- Preparation of usability test session, 27
- Pretender usability test task, 37
- Privacy. *See* Data privacy
- Problem. *See* Usability problem
- Prototype. *See* CPUX-F
- Qualitative usability test, 20
- Qualitative user survey, 73
- Quality of a usability evaluation, 8
- Quantitative usability test, 56
- Quantitative user survey, 73
- Questionnaire, 74
- Questionnaire example
 - SUS, 75
 - UMUX-LITE, 76
- Rating
 - Icons, 51
 - Of usability finding, 49
- Recognition rather than recall (heuristic), 15
- Recommendation
 - For improvement, 50
 - Usability test report, 48
- Recruiting, 23
 - Screener, 23
- Remote usability test, 62
- Report. *See* Usability test report; Usability inspection report
- Requirement. *See* CPUX-F
- Respect, 66
- Retrospective recall, 62
- RITE (Rapid Iterative Testing and Evaluation), 63
- Roles, 53
 - Usability test, 53
- Sample usability test report, 51
- Satisfaction. *See* CPUX-F
- Scenario. *See* CPUX-F, *See* Usability test task scenario
- Screener. *See* Recruiting screener
- Script. *See* Usability test script
- Selecting a usability evaluation method, 6
- Selling usability findings, 43
- Seminar, CPUX-UT, 77
- Sensitive personal information, 38
- Session. *See* Usability test session
- Severity rating, 49
- Silly usability test task, 37
- Size of usability test report, 51
- Split testing. *See* A/B testing
- Stakeholder. *See* CPUX-F
- Stakeholders, involving, 44
- Standard questionnaire, 74, 75
- Styleguide. *See* CPUX-F
- Success rate, 58
- Summary. *See* Executive summary

- Summative usability evaluation. See CPUX-F
- Survey. See User survey
- SUS, 75
- System usability scale. See SUS
- System-oriented usability test task, 38
- Tactful usability findings, 51
- Talkative moderator, 32
- Task. See CPUX-F, See Usability test task
- Task completion time, 57
- Test location, 27
- Test participant, 55
 - Confirmation to, 25
 - Ethical rules, 65
 - Incentive, 24
 - List of, 51
 - Number of test participants, 25
 - Recruiting, 23
- Test plan. See Usability test plan
- Test report. See Usability test report
- Test room, 29
- Test script. See Usability test script
- Test session. See Usability test session
- Test task. See Usability test task
- Test task scenario. See Usability test task scenario
- Think aloud, 19
- UMUX-LITE, 76
- Unmoderated usability test, 62
- Usability. See CPUX-F
- Usability evaluation, 6
 - Agile, 9
- Usability evaluation method
 - Selecting, 6
- Usability findings, 41
 - Analysis, 44
 - Communicating, 43
 - Conglomerate, 51
 - Examples, use of, 51
 - Positive, 42
 - Rating, 49
 - Selling, 43
 - Tactful, 51
 - Usability test report, 48
- Usability inspection, 12
 - Criteria, 13
 - Report, 13
- Usability lab, 29
- Usability problem, 42
- Usability test
 - Agile, 9
 - Definition, 18
 - Discount, 61
 - Ethical rules, 65
 - Location, 27
 - Plan, 21
 - Qualitative, 20
 - Quantitative, 56
 - Remote, 62
 - Report, 47
 - Report example, 51
 - Report size, 51
 - Script, 22
 - Unmoderated, 62
- Usability test participant. See Test participant
- Usability test session, 26
 - Pilot, 22
 - Preparation, 27
- Usability test task, 34
 - Closed, 35
 - Clue, 36
 - Creation of, 36
 - Derogative, 38
 - Hazy, 39
 - Imprecise, 39
 - Key task, 35
 - Open-ended, 34
 - Pretender, 37
 - Scenario, 35
- Sensitive information, 38
- Silly, 37
- System-oriented, 38
- Usability tester, 53
- User. See CPUX-F
- User control and freedom (heuristic), 14
- User experience. See CPUX-F
- User experience evaluation, 7
- User experience professional. See CPUX-F
- User group. See CPUX-F
- User interface. See CPUX-F
- User interface guideline. See CPUX-F
- User requirement. See CPUX-F
- User survey
 - Definition, 73
 - Qualitative, 73
 - Quantitative, 73
- Video summary, 52
- Visibility of system status (heuristic), 14
- Walkthrough. See Cognitive walkthrough