Usability testing methods

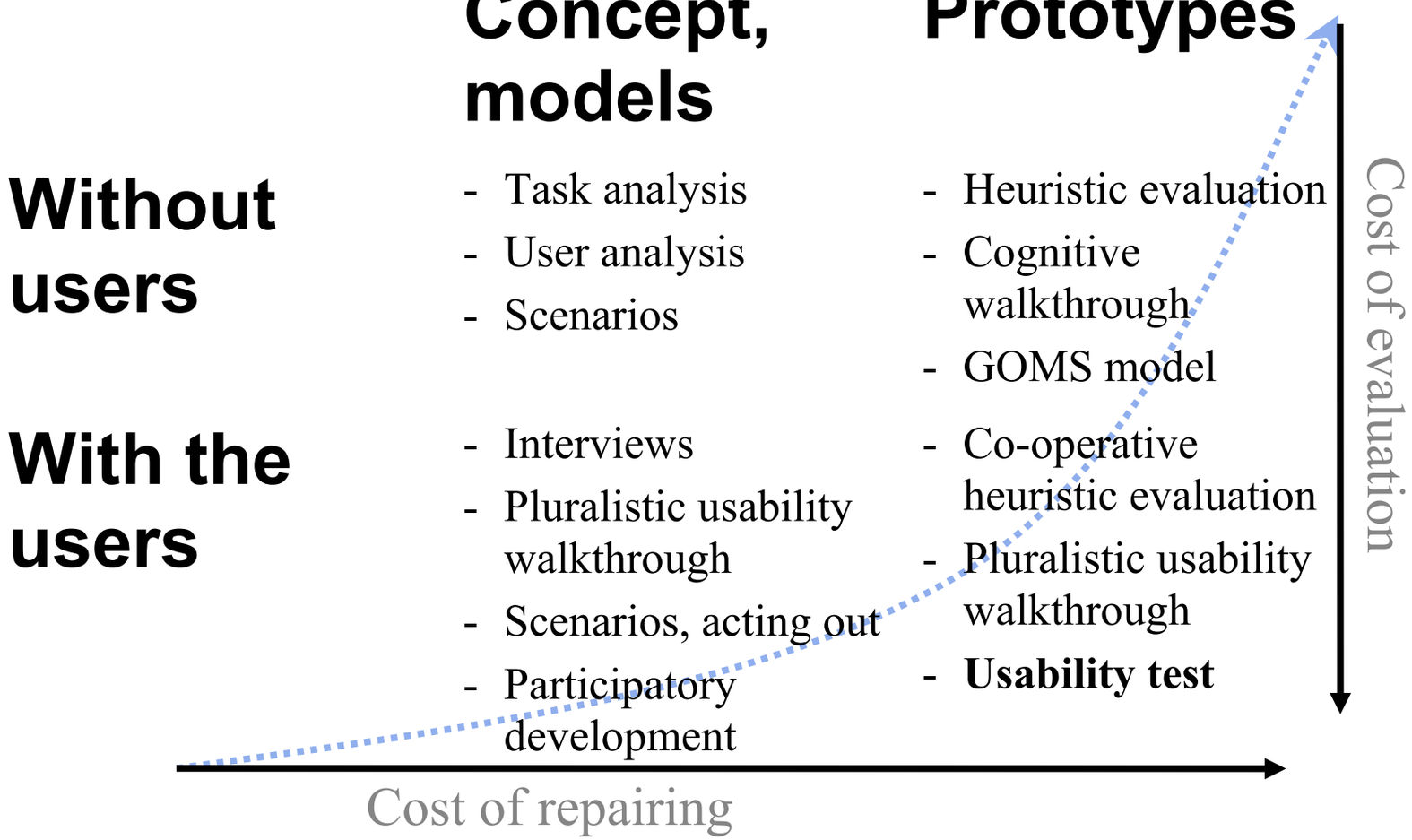
Usability Testing

- User testing with real users is the most fundamental usability method.
- User testing is (in some sense) irreplaceable, since it provides information about
  - how people use the product
  - what their problems are with the concrete interface being tested.

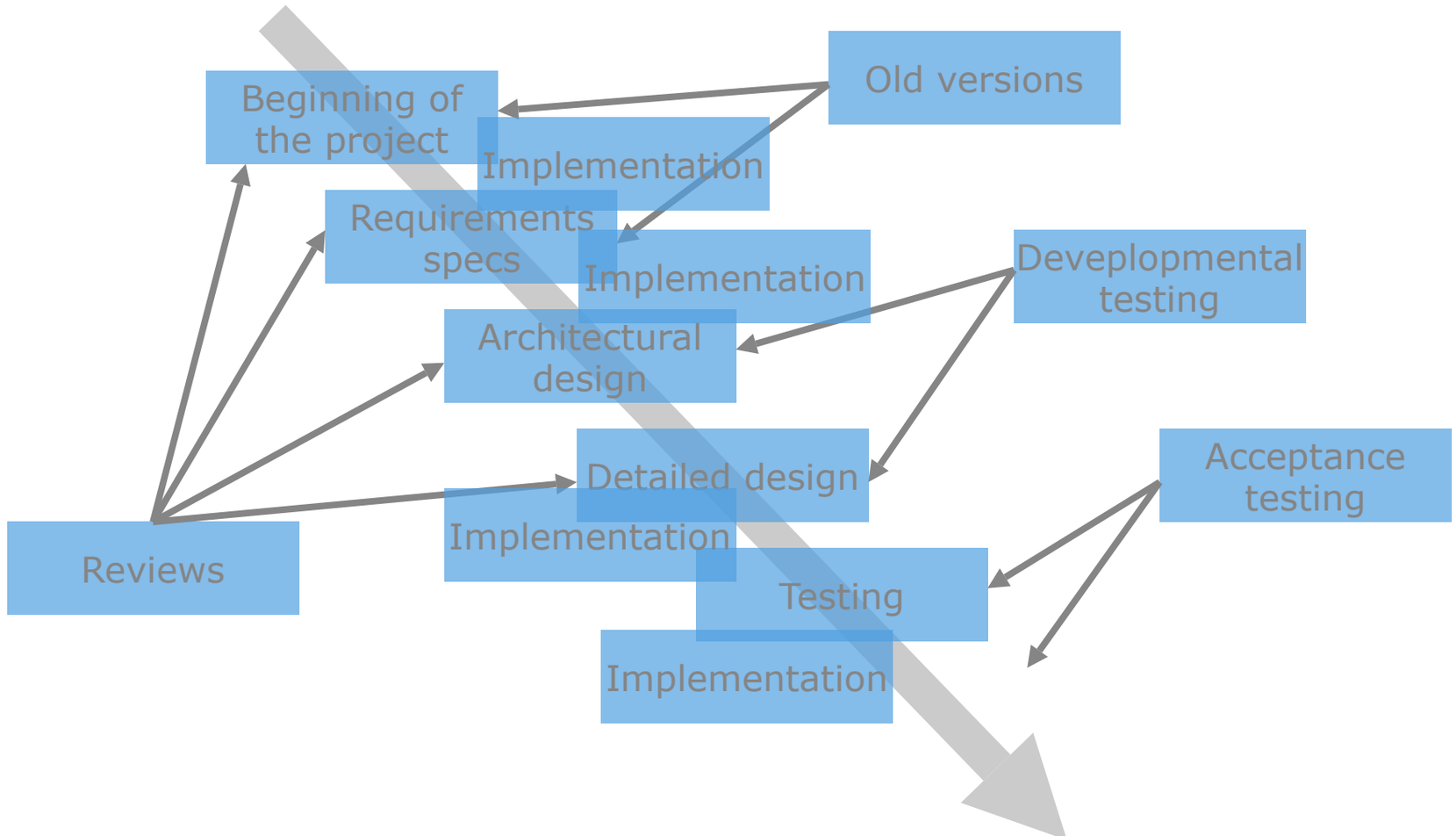
The goal

- The goal of a usability test depends on the phase of development in which it is executed.
  - **Developmental test** tries to find the best possible user interface for a product.
  - In **acceptance test** the goal is to verify that the product fulfils the usability requirements that were set.

Evaluation Methods



Different test types in different phases



Reliability

- Reliability is the question of whether one would get the same result if the test were to be repeated.
- Reliability of usability tests is a problem because of the huge individual differences between test users.
- It is not uncommon to find that the best user is 10 times as fast as the slowest user, and the best 25% of users are normally about twice as fast as the slowest 25% of the users.

Validity

- Validity is the question of whether the usability test in fact measures something of relevance to usability of real products in real use.
- Validity requires methodological understanding of the test method one is using as well as some common sense.
- Typical validity problems involve
  - using the wrong users
  - giving them the wrong tasks
  - not including time constraints and social influences.

Organising a usability test

Test Goals

- Before any testing is conducted, one should clarify the **purpose of the test** since it will have significant impact on the kind of testing to be done.
- A major distinction is whether the test is intended as a formative or summative evaluation of the UI.
- *Formative evaluation* is done in order to help improve the interface as part of an iterative design process.

Test Goals

- The main goal of formative evaluation is to learn which detailed aspects of the interface are good and bad, and how design can be improved.
- A typical method to use for formative evaluation is thinking-aloud test.
- *Summative evaluation* aims at assessing the overall quality of an interface, for example, for use in deciding between two alternatives.
- A typical method to use for summative evaluation is a "measurement" test.

Defining test goals

- What do you want? Are you looking for problem issues or are you doing measurements?
- Example goals:
  - Overall usability
  - Suitable for power users
  - Suitable for irregular use
  - Minimising need for support
  - Learnability
  - Errors

Performance Measurement

- Measurement studies form the basis of much traditional research on human factors and are also important in the usability engineering life-cycle for assessing whether usability goals have been met and for comparing competing products.
- User performance is almost always measured by having a group of test users perform a predefined set of test tasks while collecting time and error data.

Performance Measurement

- Typical quantifiable usability measurements include [Nielsen 1993]:
  - Time to complete a specific task
  - Number of tasks of various kinds that can be completed within a given time limit
  - The ratio between the successful interactions and errors
  - Time spent recovering errors
  - Number of immediately subsequent errors
  - Number of commands or other functions that were utilised by the user

Performance Measurement

- The number of commands that were never used by the user
- The number of system features that the user can remember during a debriefing after testing
- The frequency of use of the manual/help and the time spent using them
- How frequently the user manual/help solved the problem
- The proportion of positive and negative user statements
- Number of times the user express clear frustration (or clear joy)

Performance Measurement

- The number of times the user had to work around an not - solvable problem
- The amount of “dead” time when the user is not interacting with the system
- The number of times the user is side tracked from focusing on the real task
- Of course, only a subset of these measurements would be collected during any particular measurement study.

Test Plans

- A test plan should be written down before the start of the test and should address the following issues [Nielsen 1993]:
  - **The goal of the test:** what do you want to achieve?
  - Where and when will the test take place?
  - How long is each test session expected to take?
  - What computer support will be needed for the test?
  - What software needs to be ready for the test?
  - What should the state of the system be at the start of the test?
  - What should the system/network load and What should the system/network load and response times be? (not too fast or too slow);

Test Plans

- Who will serve as experimenters for the test?
- Who are the users going to be and how are we going to get hold of them?
- How many users are needed?
- What test tasks will the users be asked to perform?
- What **criteria will be used to determine** when the users have finished each of the test tasks correctly?
- What user aids will be made available to the test users (manuals, online help, etc)?
- To what extend will the experimenter be allowed to help the users during the test?
- What data is going to be collected and how will it be analysed?

Stages of a Test

- How much does the user know about the product?
- Background of the user: occupation, age etc
- A usability test can be divided into separate parts like this:
  - Describing the objectives to the user
    - The purpose is to evaluate the software, not the user
    - All information is confidential
    - All problems are important
    - The user may pause or stop the test
  - Briefing and interview

Stages of a Test

- The test itself
  - The user should have feeling of control
  - Create a scenario of what the user should do i.e., tasks
  - Tasks are usually given one at time
  - Natural tasks
  - Observe, ask, give guidance when needed
- Visual walkthrough/think aloud included
  - User says where s/he is focusing
  - How does s/he understand what s/he sees
- Debriefing, questionnaire
  - User's feeling to the software; situation; how does it relate to its competitors or previous versions; usefulness...

Getting Test Users

- Users should be as representative as possible of the intended end-users of the system.
- If more test users are to be used, they should be selected from several sub-populations to cover the main different categories of expected users.
- One should always know how much background knowledge the users have.
- Selection of the users:
- The target group analysis from the planning phase
- Criteria to find out:
  - 1) age
  - 2) education background
  - 3) previous use of computers
  - 4) experience of similar services

Choosing Experimenters

- No matter what method is chosen, somebody has to serve as the experimenter and be in charge of running the test.
- An experienced experimenter would obviously be the best choice; However, it's better to find a few usability problems than not to find any.
- The experimenter must have extensive knowledge of the application and its user interface. (System designers are often in the testing team.)

Ethical Aspects

- Even though usability test subjects are not bodily harmed, test participation can still be quite distressful experience for the users.
- Users feel a tremendous pressure to perform, even when they are told that the purpose of the study is to test the system and not the user.
- The experimenter has a responsibility to make the users feel as comfortable as possible during and after the test.

Pilot Tests

- No usability testing should be performed without first having tried out the test procedure on a few pilot subjects.
- Subjects such as one's colleagues are ok.
- During pilot testing, one will typically find that the instructions for some of the test tasks are incomprehensible to the users or that they misinterpret them.
- It is common that the tasks are more difficult than expected.

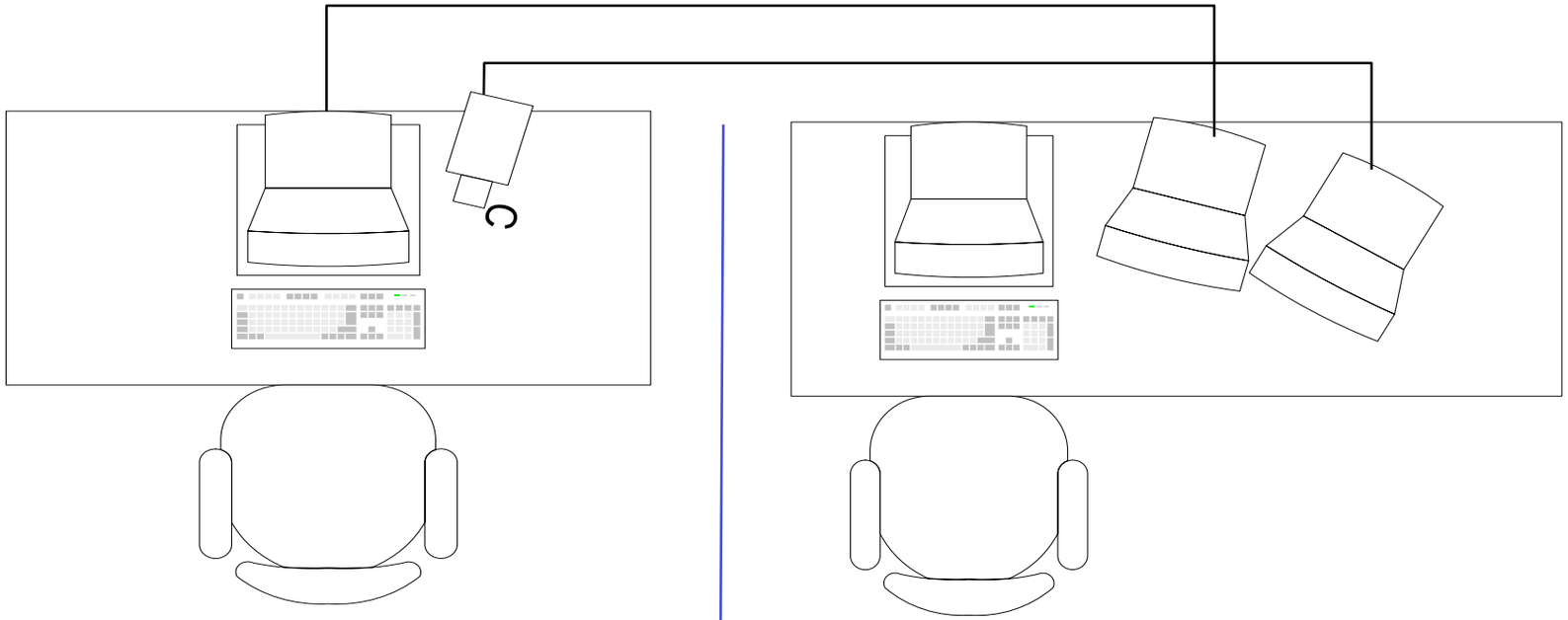
Test Tasks

- The basic rule for test tasks is to select tasks that represent real-life tasks as closely as possible.
- Also, the tasks should provide reasonable coverage of the most important parts of the UI.
- The tasks should be small enough to be completed within the time limit, but they should not be so small that they become trivial.
- All test tasks should be business-oriented and as realistic as possible.

Physical Lab Contents

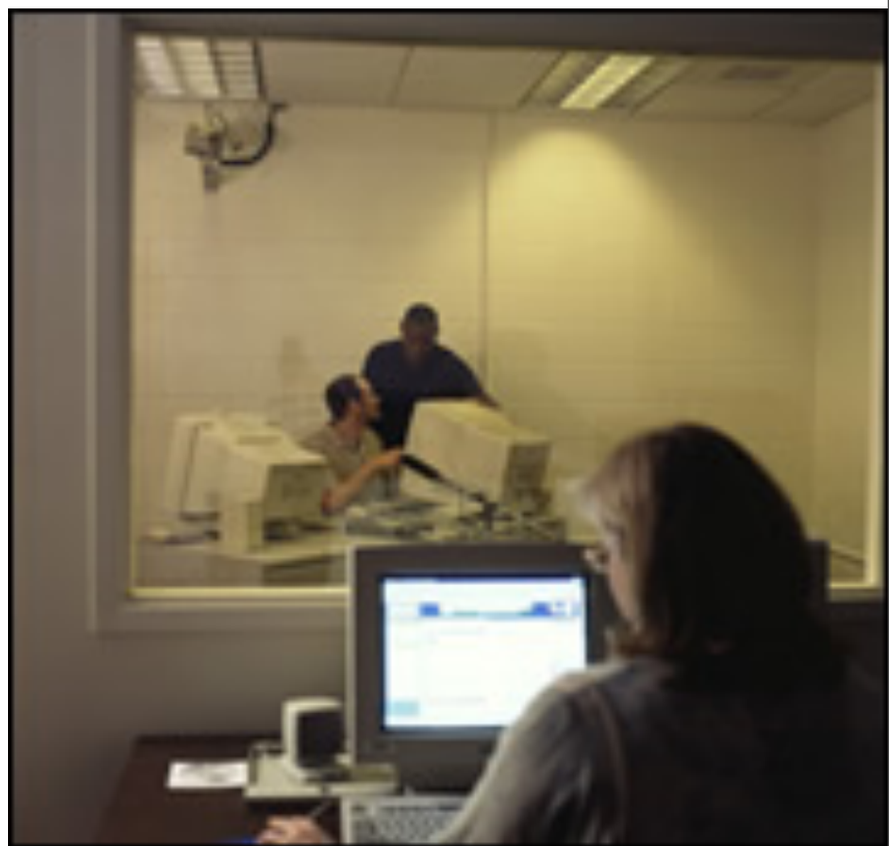
- Equipment
  - Observers' computers
  - Observation monitors
  - Participant computers
  - Cameras
  - Microphones
  - Phones
  - Headphones
  - Desks
  - Chair for interviewer
- Files/docs
  - Intro briefing script
  - Note taking template
  - Interview scripts / templates
  - Survey
- People
  - Observers
  - Participants

Usability Lab

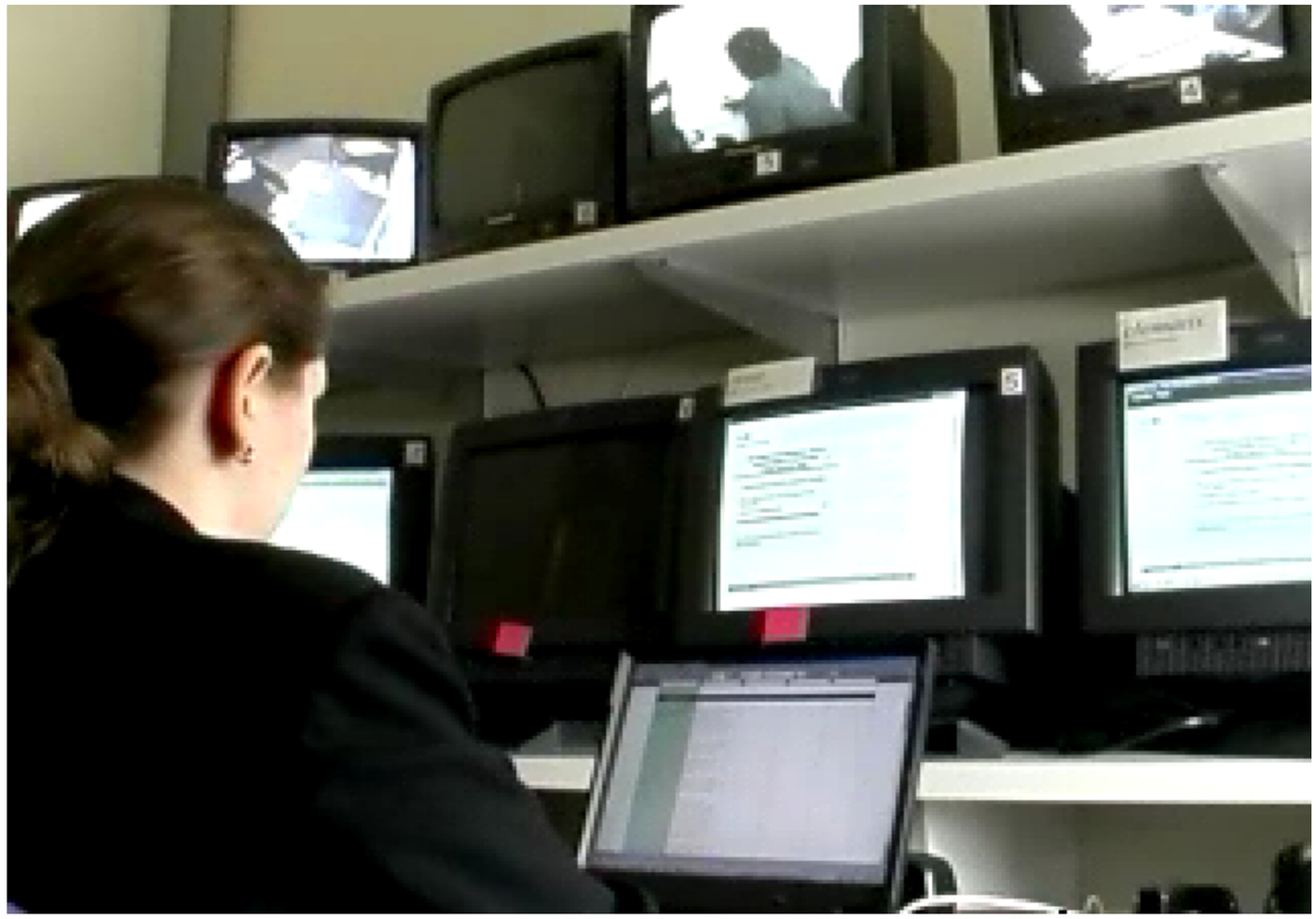


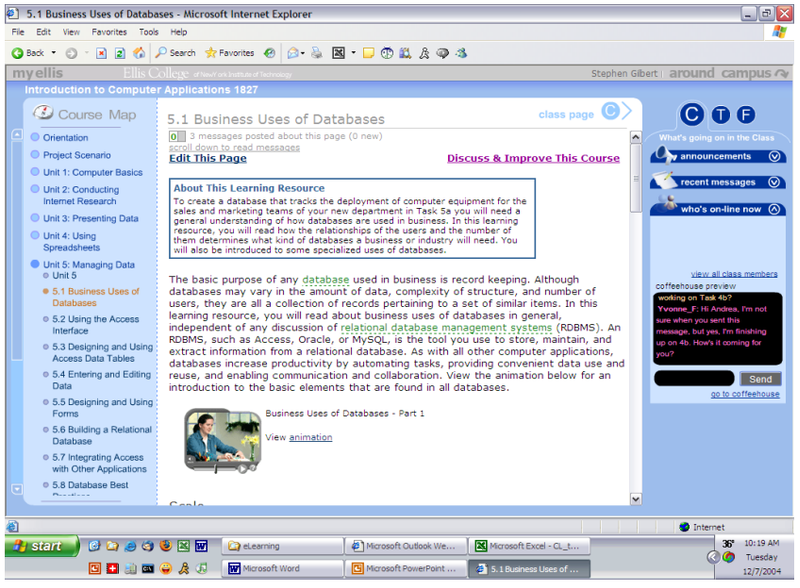
Participant

Observer



Usability Lab





Ranking the Problem

Severity:

- Cosmetic; maybe not even a problem
- A small problem that slows down learning of the use
- A problem that user can notice and learn to avoid, but causes loss of efficiency
- A big problem that is easily noticed but not easily avoided; something the user bumps into constantly
- A catastrophic problem that may cause the user to lose data or work, or gives false impression on the system quality

Ranking the Problem

- Another method (in case you are in a hurry):
  - Local bug, easy to fix
  - Logical bug, quite easy to fix
  - Fixing the problem demands redesign
- Catastrophic and easy-to-fix errors have to be fixed immediately.
- Fixing minor bugs may make conclusive changes to the overall usability.

Ranking the Problem

- When you are observing the user or analysing the test data, it would be good to note also
  - Heuristic rules
    - At least as a reminder for yourself (development team)
    - Easier to communicate and correct
  - Pervasiveness
    - Does the problem exist everywhere in the software (and in every version of it)
  - Prediction on how often user “finds” the problem

Test report

- Normal usability test report includes the following chapters:
  - Description of the product and its normal use
  - Short description of the method used and test users
  - Functions that were tested; test scenarios and tasks
  - Result of the test (analysis and measurements)
  - Conclusion of the product and test by the testing manager
- 20-200 pages, depends on the testing method, number users etc.

Test report

- Write good explanations – why something should be fixed? The developers will make the changes if they understand the need for them.
- Do not blame the designers!
- List also good features in order to prevent them from being ”fixed”.
- Have a meeting with designers to make sure that they understand the findings that were reported to them.

Testing methods

Thinking Aloud

- Thinking aloud may be the single most valuable usability engineering technique – often used in other methods also - Mixing methods
- A thinking-aloud test involves having a test subject use the system while continuously thinking out loud.
- By verbalising their thoughts, the test users enable us to understand how they view the computer system, and this makes it easier to identify the users' major misconceptions.

Thinking Aloud

- We do not know what people think; thinking aloud gives a glimpse of what is going on.
- Thinking aloud in this manner involves tasks
- Users are asked to ”describe what they are doing”.

- It’s usually necessary to explain the method to the user in terms of: ”think aloud; tell [me] what you are doing, what you are looking at or looking for, what do you expect. And always when you’d like to know something, please ask.”
- **First task is usually quite simply** to help the user to concentrate on thinking aloud.

Thinking Aloud

- Help the user to think aloud during testing to get better results (and more feedback).
  - Questions such as ”What are you looking for now”, ”what do you think about what you see” may help the user to think aloud.
- If the user ceases thinking aloud:
  - There may be something worth of reading, or
  - The user has to concentrate on the interface.

Usability Test Variations

- Testing in pairs
  - Two users at a time; more natural discussion and problem solving
- **Observation as testing**
  - Users are observed as they do their normal tasks

Usability Test Variations

- Freeform walkthrough
  - User tries out the product without guidance or task
- Pluralistic usability walkthrough
  - Typically 2-6 users and 1-3 developers at the same time

Are user manuals allowed?

- If users are allowed to have user's manuals in the test, everything changes. Then you'll be testing also the manual.
- User's manual also makes it difficult to analyse the data, since some users benefit more of the manuals than others.
- Also a manual that users read before the test will change the results since some people learn better by reading a book.

References

- Jakob Nielsen: Usability Engineering, 1993
- Irmeli Sinkkonen, Hannu Kuoppala, Jarmo Parkkinen, Raino Vastamäki: Psychology of Usability, 2006
- Usability.net

Readings & webpages

- **Links to different kind of templates needed in "application/service"** production:
  Usability.gov .: http://www.usability.gov/templates/
- **Example of technical testing templates:**
  http://www.klariti.com/templates/Test-Plan-Template.shtml
  http://www.developsense.com/testing/TestPlanOutline.doc
- **Examples of different kinds of guidelines:**
  http://www.serco.com/usability/research/guidelines/index.asp
- **Graphical user interface guidelines examples:**
  Mac full detailed guideline
  http://developer.apple.com/documentation/UserExperience/Conceptual/OSXHIGuidelines/index.html
- **Accessible Web Design**
  http://www.rnib.org.uk/digital/hints.htm
- Nielsen J. 1993. Usability Engineering, Academic Press Limited, London
- Nielsen J.2002/1994. Jacob Nielsen's Web site for Usable Information Technology.
- **The Alertbox.**
  http://www.useit.com/alertbox/ &
  http://www.useit.com/papers/heuristic/heuristic_list.html.
- Shackel, B. 1991. Usability – context, framework, design and evaluation. In Shackel, B. and Richardson, S. (eds.). Human Factors for Informatics Usability. Cambridge University Press, Cambridge.
- **WWW-document.**
  http://jthom.best.vwh.net/usability/
- **Web Accessibility Initiative**
  WWW-document. http://www.w3.org/WAI/