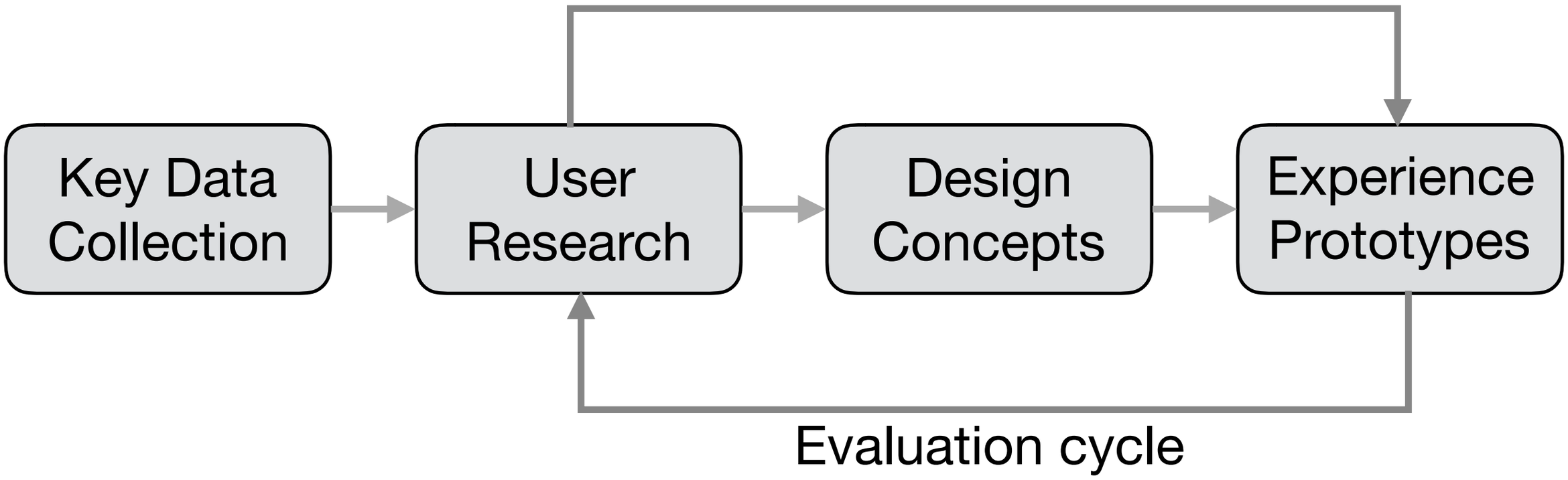
INF03 Expérience Utilisateur

7. Evaluation and tests – 2 Analytical & Empirical methods

Aurélien Tabard

Rappels des dernières séances



Evaluation and tests

Last week :

- Introduction
- Approaches to evaluation

Today :

- **Analytical methods**
- **Empirical methods**

Next week :

- Experimental design
- Evaluation 2.0, scaling up

Two types of analytical methods

- Introduction
- Approaches to evaluation

- **Analytical methods**
  - **Based on models**
  - Based on inspection
- Empirical methods

- Experimental design
- Evaluation 2.0, scaling up

Model based evaluation

Popular in the 80's

Work for low level tasks, e.g., command selection

Examples

- GOMS (Goals, Operators, Methods, and Selection rules)
- KLM

Regaining popularity today with big data

KLM

Analyze a task and split-it into atomic actions

Estimate the duration of each step based on a task model

Estimate the time of the whole task

> Enables to optimize design before implementation

From http://www.cs.umd.edu/class/fall2002/cmsc838s/tichi/printer/goms.html

KLM

Description	Operation	Time (sec)
Reach for mouse	H[mouse]	0.40
Move pointer to "Replace" button	P[menu item]	1.10
Click on "Replace" command	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Specify word to be replaced	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Point to correct field	P[field]	1.10
Click on field	K[mouse]	0.20
Home on keyboard	H[keyboard]	0.40
Type new word	M4K[word]	2.15
Reach for mouse	H[mouse]	0.40
Move pointer on Replace-all	P[replace-all]	1.10
Click on field	K[mouse]	0.20
Total		10.2

Limits

- Predictions only valid for expert users not making errors
  - Experts are also error-prone
  - doesn’t take into account novice and intermediate users
  - il existe des extensions qui essaient de modéliser l’apprentissage
- Tasks should have a clear goal
  - Many tasks are not as directed especially when designing for experience rather than efficiency.
- Doesn’t take into accounts individual differences
  - Based on means
- Doesn’t take into account social and organizational aspects of tasks
- No information of quality of the experience
- Not representative of modern theories of human cognition
  - Supposes a linear model of cognition with only one ongoing activity

Two types of analytical methods

- Introduction
- Approaches to evaluation

- **Analytical methods**
  - Based on models
  - **Based on inspection**
- Empirical methods

- Experimental design
- Evaluation 2.0, scaling up

Inspections et expert reviews

Like we did last week

- Throughout the development process
- Driven by designers, developers, usability experts
- Useful to identify problems
- From one hour to a week of work
- Prefer structured approaches
- Solutions can be suggested

Pros and cons of expert reviews

- Results of informal reviews and inspections are often directly used to change the product
  - ... still state of the art in many companies!
  - The personal view of the CEO, or his partner ...
- Really helpful evaluation
  - Is explicit
  - Has clearly documented findings
  - Can increase the quality significantly
- Expert reviews and inspections are a starting point for change

Usability guidelines (les critères ergonomiques)

- Don Norman's principles:
  - visibility, affordances, natural mapping, and feedback
- Ben Shneiderman's 8 Golden Rules of UI design
- Bruce Tognazzini's 16 principles:
  - http://www.asktog.com/basics/firstPrinciples.html
- Christian Bastien's Ergonomic Criteria
- Jakob Nielsen's Heuristics

Heuristic evaluation

Designed as a discount evaluation method based on inspection:

- Quick and cheap UI evaluation
- http://www.useit.com/papers/heuristic/

Principles:

- There is a list of important rules for UIs: “heuristics”
- Heuristics can be checked by experts with an precise outcome

10 Usability Heuristics

http://www.useit.com/jakob/photos/

- Meet expectations
  1. Match the real world
  2. Consistency & standards
  3. Help & documentation
- User is boss
  4. User control & freedom
  5. Visibility of system status
  6. Flexibility & efficiency
- Errors
  7. Error prevention
  8. Recognition, not recall
  9. Error reporting, diagnosis, and recovery
- Keep it simple
  10. Aesthetic & minimalist design

Evaluation and tests

Last week :

- Introduction
- Approaches to evaluation

Today :

- Analytical methods
- **Empirical methods**

Next week :

- Experimental design
- Evaluation 2.0, scaling up

Why use empirical methods?

It is difficult to know the quality of an experience until people tried your product.

Limits of expert methods:

- Experts know too much
- Experts know too little (e.g. on users' tasks)
- Hard to predict what "real" people will do

Why use empirical methods?

Identify usability problems

Gather data about the performance of the product

Evaluate user satisfaction when using the product

Empirical methods

- Focus groups
- Field studies
- Usability studies
- Physiological measures
- Controlled experiments

Question

What kind of method should be used

when introducing a new keyboard for iOS ?

- Focus groups
- Field studies
- Usability studies
- Physiological measures
- Controlled experiments

Principles shared by all empirical studies

- Focus groups
- Field studies
- Usability studies
- Physiological measures
- Controlled experiments

Recruiting participants

Representative users

- know about the subject matter
- mastery of related tasks

Get as close as possible

- system for medical doctors
  -> recruit medical students

Use “carrots” to get participants

Ethics!

Pressures on participants:

- Performance anxiety
- Feels like an intelligence test
- Comparing self with other subjects
- Feeling stupid in front of observers
- Competing with other subjects

Respect et control

- Don’t waste time
- Confort
- Informed consent
- Privacy
  - Data is collected anonymously and stays private
- Control
  - User can give up whenever they want

Before the test

- Time
  - Pilot-test all materials and tasks
- Comfort
  - “We’re testing the system; were not testing you.”
- Privacy
  - Your test results will be completely confidential.
- Information
  - Brief about purpose of study
  - Inform about recording & observers
  - Answer any questions beforehand (unless biasing)
- Control
  - You can stop at any time.

During the test

- Time
  - Eliminate unnecessary tasks
- Comfort
  - Calm, relaxed atmosphere, take breaks in long session
  - Never act disappointed
  - Give tasks one at a time
  - First task should be easy, for an early success experience
- Privacy
  - Users’ boss shouldn’t be watching
- Information
  - Answer questions (again,where they won’t bias)
- Control
  - Users can quit entirely

After the test

- Comfort
  - Say what they've helped you do
- Information
  - Answer questions that you had to defer to avoid biasing the experiment
- Privacy
  - Don't publish user-identifying information
  - Don't show video or audio without users permission

Empirical methods

- **Field studies**
- Usability studies
- Physiological measures
- Controlled experiments

Field studies

Benefits:

- Better understanding of acceptance
- Longer term, good to understand learning dynamics, collaboration, integration in everyday activities, etc.

Problems:

- Expensive
- Requires a reliable product (or prototype)
- Cumbersome gathering of observations

Empirical methods

- Field studies
- **Usability studies**
- Physiological measures
- Controlled experiments

What is usability testing?

Usability testing is a means for measuring how well people can use some human-made object (such as a web page, a computer interface, a document, or a device) for its intended purpose, i.e. usability testing measures the usability of the object.

Metrics

Ease of learning

- learning time, …

Ease of use

- completion time, error rate, …

User satisfaction

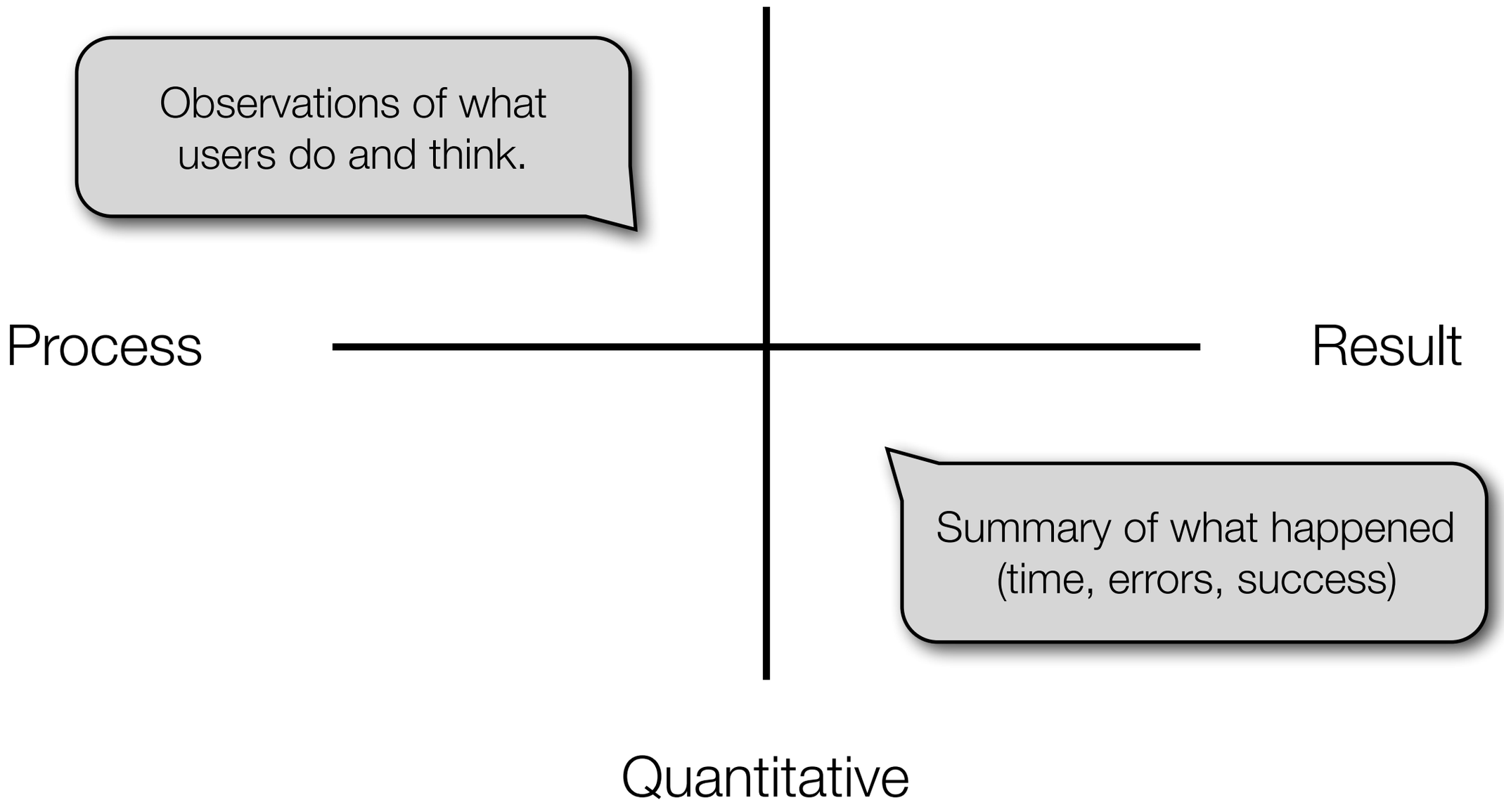
- survey…

User experience

- Attrackdiff

~~**intuitive**~~ ~~**natural**~~

What kind of data to record?



What to record: quantitative data

- **Success rates**
- **Accuracy / Error rates**: How many mistakes did people make? And were they fatal or recoverable with the right information?
- **Time on Task**: How long does it take people to complete basic tasks? (For example, find something to buy, create a new account, and order the item.)
- **Pages visited**, number of steps to reach goal...
- **Recall**: How much does the person remember afterwards or after periods of non-use?
- **Emotional Response**: Ratings on a satisfaction questionnaire, How does the person felt about the tasks completed? (Confident? Stressed? Would the user recommend this system to a friend?)

What to record: qualitative

- How people reacted to the system.
- How participants understood it.
- Which pathways participants took.
- Which problems participants had (critical incidents).
- What participants said as they worked.
- Participants' answers to open-ended questions.

You need a plan!

A good plan for usability testing gives the participants:

- a goal/task (what to do or what question to find the answer for)
- data, if needed, that a real user would have when going to the site to do that task

You can give the scenario as just the statement of the goal/task or you can elaborate it a little with a very short story that adds motivation to get to the goal.

Participants

The participants must be like the people who will use your product.

Be ready to screen participants (do not grab the first person in the corridor)

Plan on a cost associated with finding the people

- you may still need to plan on incentives to get participants to participate ...

Test!

Make sure you have everything you need

- the prototype you are going to test
- the computer set up specified in your test plan
- consent forms
- questionnaires, if you are using any
- the participant's copy of the scenarios
- cameras, microphones, or other recording equipment
- folders to keep each person's paperwork in if using paper

Do a dry-run and a pilot test

Before starting

- You should know, and have written down
  - objective
  - description of system being testing
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures
- Will help you design a good usability test
- Will help you figure out how to analyze your data

Usability laboratory

- Dedicated testing room
  - Instrumented with data collection e.g. mic., cameras, loggers…
- Separate observation room
  - Usually connected to testing room
  - One-way mirror and audio system
- Test users on scenarios
  - “Think aloud” technique
  - Decide whether to interrupt or not
  - Keep variances among tests low

From C|Net “How Google tested Google Instant”
http://news.cnet.com/8301-30684_3-20019652-265.html

Think aloud

Need to know what users are thinking, not just what they do

Ask users to talk while performing tasks

- tell us what they are thinking
- tell us what they are trying to do
- tell us questions that arise as they work
- tell us things they read

Prompt users **very** often

- *“tell me what you are thinking”*
- keep track of anything you do give help on

Usability testing analysis and limitations

- Summarize the data
  - make a list of all critical incidents
  - include references back to original data
  - try to judge why each difficulty occurred
- What does data tell you?
  - UI work the way you thought it would? users take approaches you expected?
  - something missing?
- Update task analysis & rethink design
  - rate severity & ease of fixing incidents
  - fix both severe problems & easy fixes

Usability measures

Situations in which numbers are useful

- time requirements for task completion
- successful task completion
- compare two designs on speed or # of errors

Measures

- time is easy to record
- errors or successful completion is harder
  define in advance what these mean

Do not combine efficiency measures with thinking-aloud.

- talking can affect speed & accuracy

Usability lab on the cheap

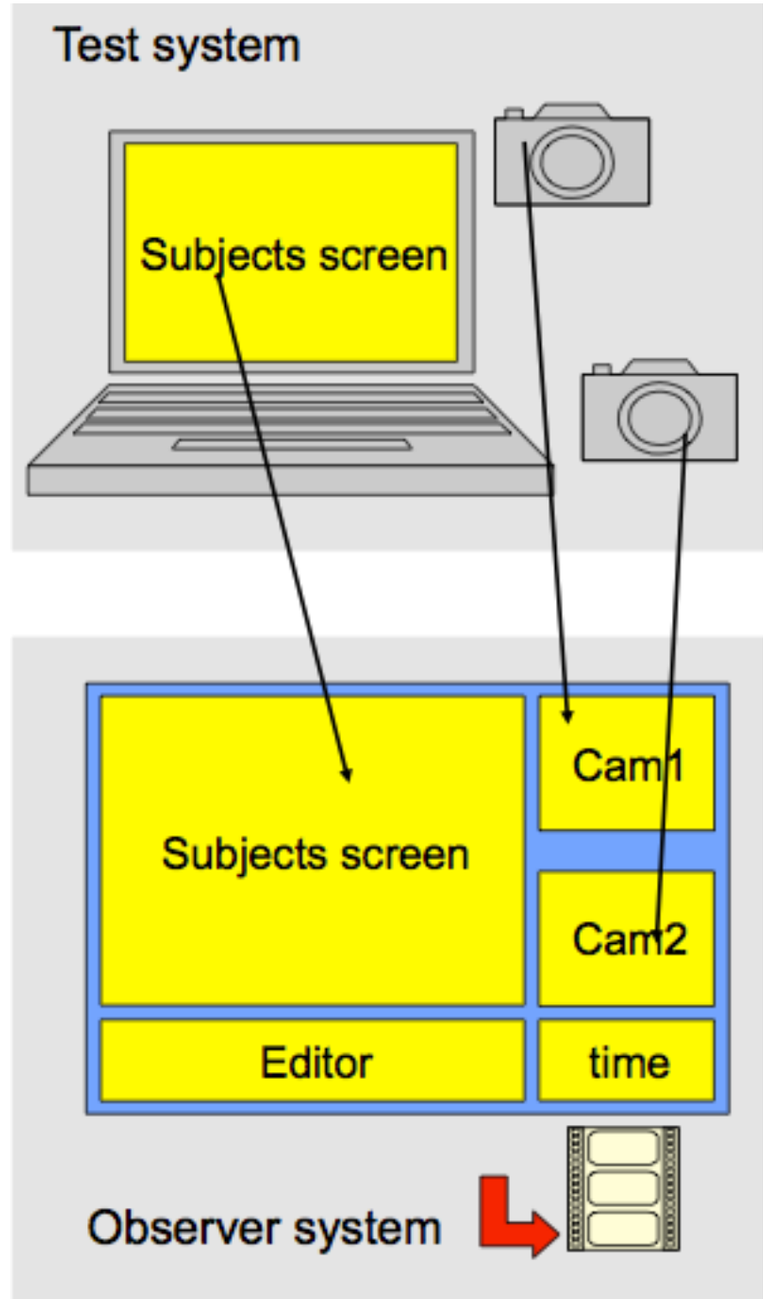
Goal: Integrate multiple views

- Capture screen with pointer
- View of the person interacting with the system
- View of the environment

Setup:

- Computer for the test user
- Computer for the observer

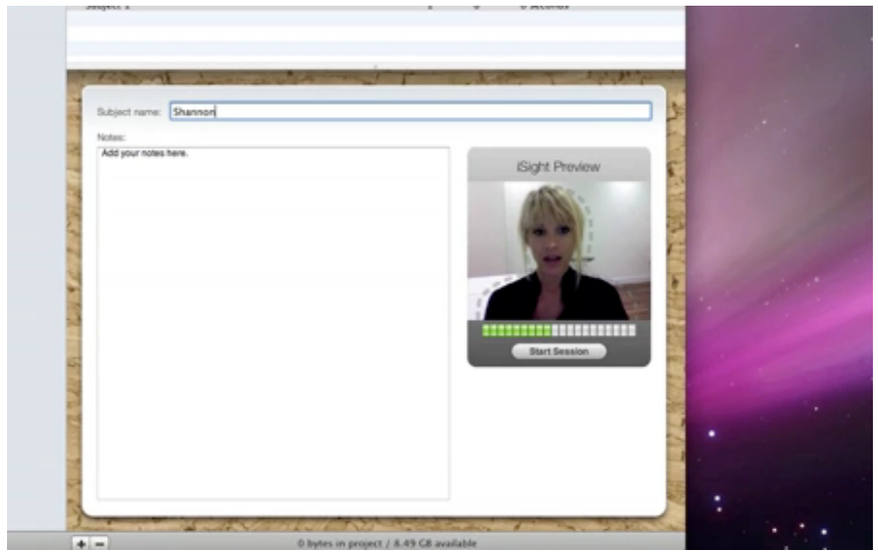
Debrief with the users afterwards



Existing tools

"New" ones:

- Lookback https://lookback.io/features/in-person/
- UXCam https://uxcam.com/
- Silverback http://silverbackapp.com/



Old school:

- Morae http://www.techsmith.com/morae.html
- Ovo studio http://www.ovostudios.com

Quick and cheap

Empirical methods

- Field studies
- Usability studies
- **Physiological measures**
- Controlled experiments

Physiological measures

Attention

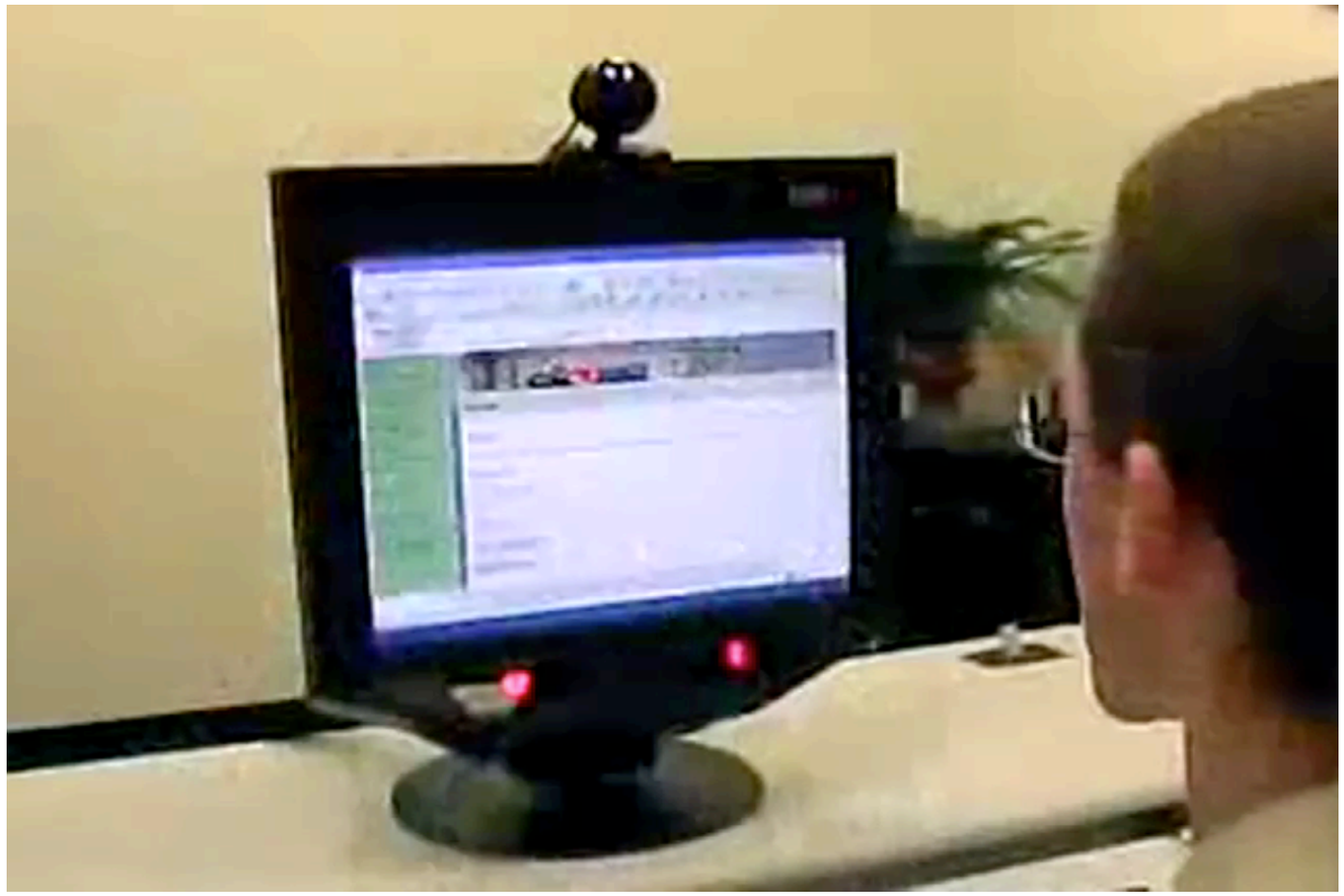
- Eye tracking widespread, robust, getting cheaper

Stress

- skin conductivity

Brain activity

- still experimental



Eye-tracker - © Kent State University (US)

Empirical methods

- Field studies
- Usability studies
- Physiological measures
- **Controlled experiments, aka A/B testing.**
  - Next week