

Evaluation

Chapter 8



Evaluation

- It is about assessing the usability of a system.
- In HCI we evaluate interfaces and systems to:
 - Determine how usable they are for different user groups
 - Identify good and bad features to inform future design
 - Compare design choices to assist us in making decisions
 - Observe the effects of specific interfaces on users

The Importance of Evaluation

- To ensure that people **can use** the system/product, so that they will **like it**
- To allow designers to concentrate on **real problems** rather than imaginary one
- To provide a means for suggesting improvements
- To ensure that problems are fixed before the system/product is shipped out.

Goals of Evaluation

- Assess extent of system functionality
 - it must match with the user's requirements.
 - matching the use of the system to the user's expectations of the task
- Assess effect of interface on user (usability)
 - This includes considering aspects such as how easy the system is to learn, its usability and the user's satisfaction with it.
 - enjoyment and emotional response (application for leisure & entertainment purpose)
- Identify specific problems
 - At the aspects of the design which, when used in their intended context, cause unexpected results, or confusion amongst users.

- Why? -To validate users' requirements to see if they can use it and whether they will like it.
- What?- Early designs(Lo-fi prototype) and a complete design(Hi-fi prototype).
- Where?- On field or in laboratory.
- When?- Throughout the HCI iteration process.

Types of Evaluation

- Formative evaluation
 - Evaluation done during the design phase. It can be done by the design team or by involving real users.
- Summative Evaluation
 - Evaluation performed with finished product. It is mostly done by external users.



Types of Methods

- Two broad headings: expert analysis and user participation.
- Expert analysis:
 - The basic intention is to identify any areas that are likely to cause difficulties
 - Can be used at any stage in the development process from a design specification, through storyboards and prototypes, to full implementations
 - Relatively cheap, since they do not require user involvement.
 - However, they do not assess actual use of the system, only whether or not a system upholds accepted usability principles.
- User participation in evaluation tends to occur in the later stages of development when there is at least a working prototype of the system in place.



Evaluating Through User Participation

Styles of Evaluation

- Laboratory studies
- Field Studies



Styles of Evaluation

Laboratory studies

Users are taken out of their normal work environment to take part in controlled tests, often in a specialist usability laboratory

- Advantages:
 - specialist equipment available
 - uninterrupted environment
- Disadvantages:
 - lack of context
 - difficult to observe several users cooperating



Appropriate:

- if system location is dangerous
- very constrained single-user tasks may be adequately performed in a laboratory
- compare alternative designs within a controlled context
- deliberately want to manipulate the context in order to uncover problems or observe less used procedures

- One of the most **expensive, time consuming** and **complex** ways to evaluate the UI of a system
- To carried out such an evaluation will **require**:
 - A **well-equipped laboratory** (usability lab - fitted with video monitoring equipments)
 - Example, one camera might record user's activities at the interface, keyboard & mouse and another might focus at the user
 - **Voluntary participants** (i.e. testers to complete a list of tasks)
 - **A group of observer** (to look at the participant doing the work behind a one-way mirror)

Guidelines for conducting usability testing in lab.

Before the evaluation begin

- Establish objectives and information requirements
- **Pre-test everything** before the actual test takes place
- Make the users feel **comfortable**
- **Emphasize** evaluation of the system not the user
- **Acknowledge** that the software may have usability problems
- Let users know that they **can stop at any time**
- **Explain** monitoring equipments used in the lab
- Gain co-operation with contact
 - Inform user the reason of your study and the observation technique you intend to carry out.
 - Obtain user's consent.
 - Inform user that individual results will be kept completely **confidential**.



Guidelines for conducting usability testing in lab.

During the evaluation

- **Never** ask the user perform unnecessary tasks
- Keep a **relaxed** atmosphere in the lab.
- Try to be unobstructive. Do not let yourself or equipment get in the way.
- Hand out test tasks one at a time (don't overwhelm)
- **Never** indicate displeasure with the user's performance
- **Stop** the evaluation if it becomes too unpleasant

Guidelines for conducting usability testing in lab.

After the evaluation

- Answer any questions the user may have
- Analyse, summarise, and report in relation to the objectives set out at the beginning.



Styles of Evaluation

Field Studies

- Advantages:
 - natural environment
 - context retained
- Disadvantages:
 - distractions
 - noise
- Appropriate
 - where context is crucial.



Observation + Think Aloud

- Observation: user observed performing task
- Think Aloud: user asked to describe what he is doing and why, what he thinks is happening etc.
- Advantages
 - simplicity - requires little expertise
 - can provide useful insight
 - can show how system is actually used
- Disadvantages
 - subjective, selective (user may think before describing)
 - act of describing may alter task performance



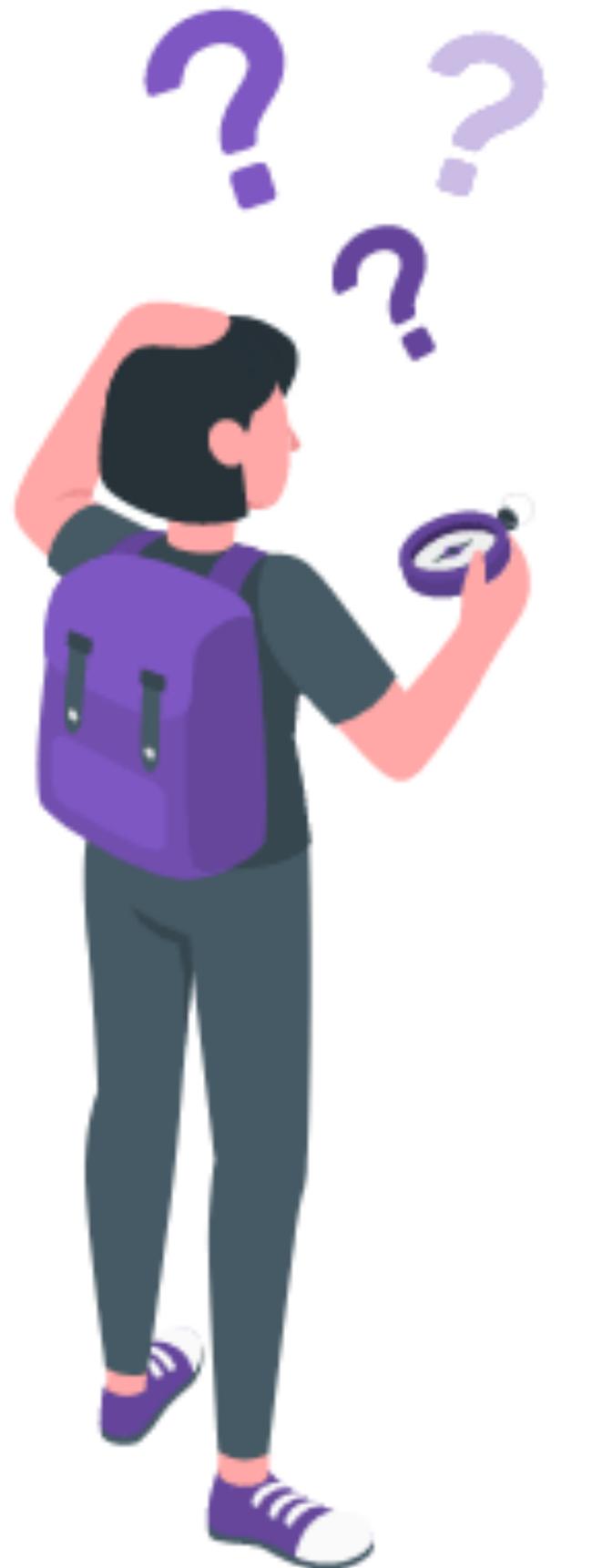
- **Useful** for gathering info such as:
 - What facilities are **frequently used**
 - How much **time** is needed to complete a task
 - Which interfaces are particularly **confusing and frustrating**
 - How users **react** to error messages
- Be aware that people may alter their way of doing things or change their performance level. This phenomenon is called the Hawthorne Effect (Observer Effect)

- Avoiding/minimising Observer Effect:
 - **Minimise** distractions
 - **Make it clear** that you are **not** evaluating the their personal performance instead the actual subject under evaluation is the UI.

During observation, record what you have observed,examples:

- The amount of **time** needed to complete a task
- Interfaces that are particularly **confusing and frustrating**
- Users **reaction** to error messages

Decide on the recording technique you will use. (Hand-written, notes, Audio, Video). The more complete your record, the longer it takes to analyse.



How to collect and record feedback from respondents?

Query Techniques

- Interviews
- Questionnaires



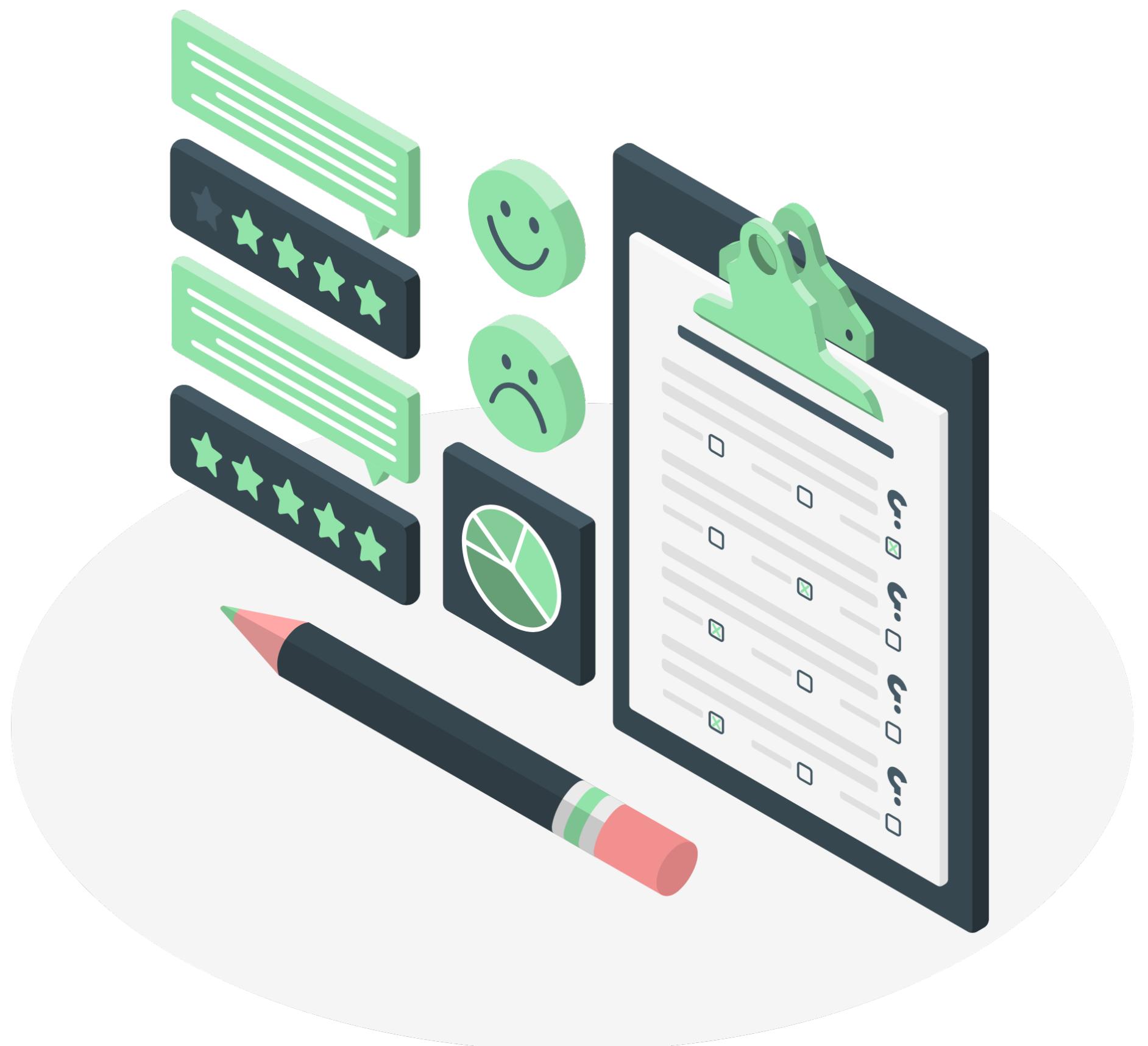
Query Techniques

Interviews

- analyst questions user on one-to-one basis (face to face or through telephone)
- Advantages
 - can be varied to suit context(change some questions according to user's reaction)
 - issues can be explored more fully
 - can elicit user views and identify unanticipated problems
- Disadvantages
 - interviewee can be very subjective and often **reluctant** to criticise openly
 - time consuming.

Questionnaires

- Set of fixed questions given to users
- Advantages
 - quick and reaches large user group
 - can be analyzed more systematic
- Disadvantages
 - less flexible
 - less probing
- Require skill to produce a **good questionnaire**
- Poor returned rates



In HCI, questionnaires have long been used to evaluate UIs.

Some notable questionnaires:

Questionnaires	Inst	Year
Questionnaire for UI Satisfaction (QUIS)	Maryland	1988
Perceived usefulness & ease of use (PUEU)	IBM	1989
Purdue Usability Testing Questionnaire (PUTQ)	Purdue	1997
Computer System Usability Questionnaire (CSUQ)	IBM	1995
System Usability Scale (SUS)	Redhatch	1986

Questionnaires

Question Format

- Closed questions, Example:

Which aspects of the interface did you like most?

- A. the color B. the sound C. the ease of moving from area to area

- Advantage
 - **Easy to be answered** because answer can be chosen from a given list
- Disadvantage
 - Could distort findings because it might suggest things to people that may not otherwise occur to them

Questionnaires

Question Format

- Open questions, Example:

What aspects of the interface did you like most?

- Advantage
 - Respondents can answer the question **in any way they like**
- Disadvantage
 - Obtained rich source of data but they are **difficult to analyzed**

Questionnaires

Question Format

What factors are important to you when buying a software product?

- Price
- Usability
- Features
- Support

Checklists

What is your favorite pizza topping?

- Pepperoni
- Mushrooms
- Anchovies
- Sausage
- Artichoke hearts

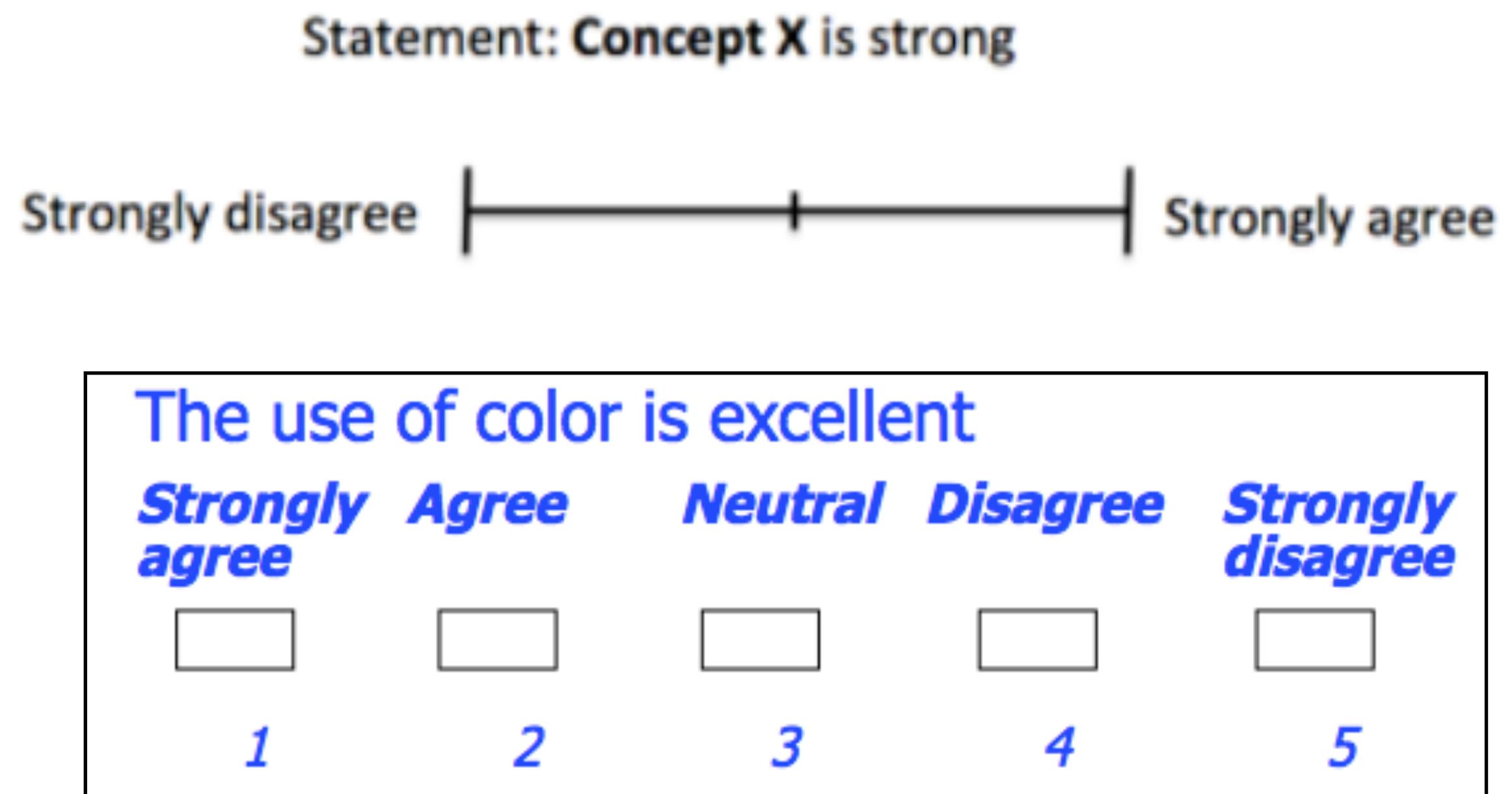
Multiple-choice questions

Questionnaires

Question Format

Scalar questions

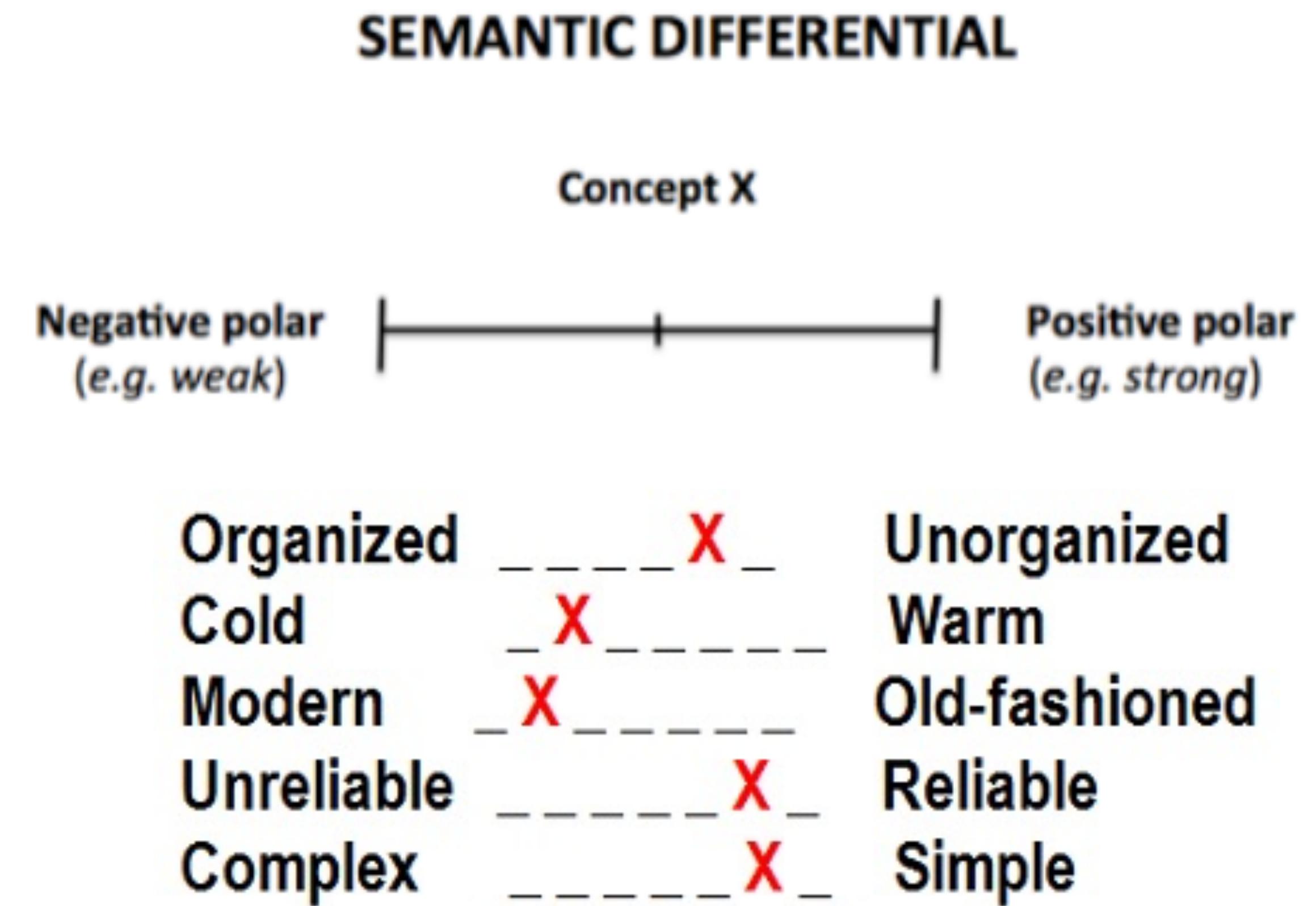
- Likert scales
 - A range of values describing how much a person agree or disagree with a particular statement.,E.g. Strongly disagree vs Strongly agree
 - Hard to quantify anything other than having participants to agree/disagree to a question/statement.



Questionnaires

Question Format

- Semantic differential scales
 - Used to measure meaning of things and concepts.
 - Measures connotative meaning – which is the suggestive significance of a word. E.g. good vs bad, sad vs happy, important vs unimportant, expensive vs inexpensive, quick vs slow.



Questionnaire Design Issues

- Can the question be easily understood?
- Is the question too vague or too precise?
- Is the question biased?
 - Eg. “You support Arsenal, don’t you?” or “Are you a criminal?”
- Is the question necessary to the evaluation?
- Will respondents be willing to provide the information?
- Is the question applicable to all respondents?

Recommendation for Questionnaire Design

- **Closed questions is preferred over open questions** - Whenever, possible, ask closed question and offer a range of possible answers.
- **Questions need to be structured carefully** – begin with general/easy questions, detailed/difficult questions last
- ~~Questionnaire must not be too long (2 side of A4 is quite enough)~~ - Long questionnaire may deter participation

Steps in conducting a Survey by Questionnaires

In general, the steps can be as follows:

- Decide on the target population and sample size
- **who** do you want to ask and **how many**
- Develop the **questions**
- **Pre-test** the questionnaire
- **Conduct** the survey
- Collect data and Analyze data collected

Experimental Technique

- an aspect of a design is to use a controlled experiment.
- evaluator chooses hypothesis to be tested
- a number of experimental conditions are considered which differ only in the value of some controlled variable.
- changes in behavioural measure are attributed to different conditions

Experimental Technique

Experimental design

- **within groups design**
 - each subject performs experiment under each condition.
 - transfer of learning possible
 - less costly and less likely to suffer from user variation.
- **between groups design**
 - each subject performs under only one condition
 - no transfer of learning
 - more users required
 - variation can bias results.

Experimental Technique

A/B testing

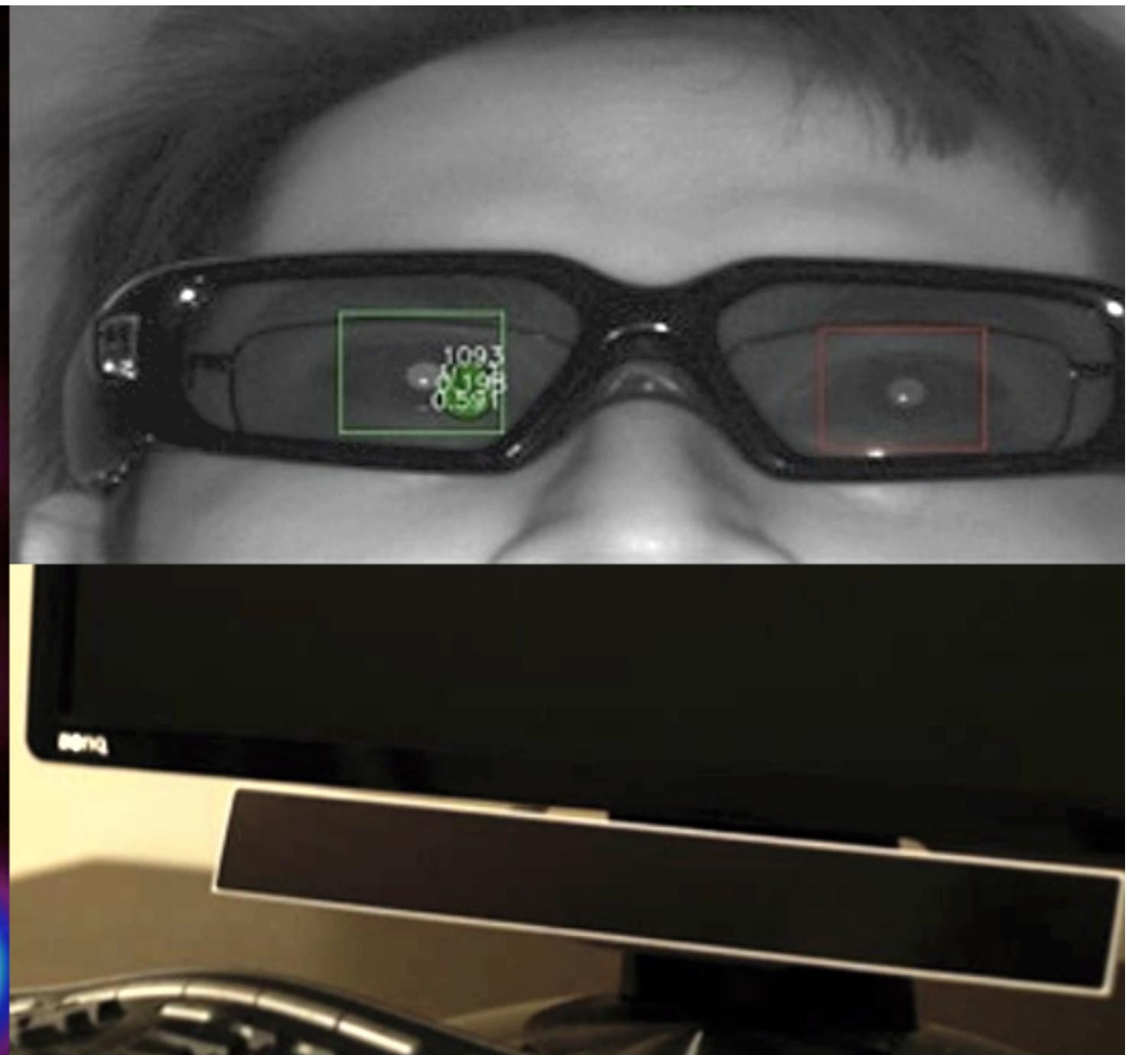
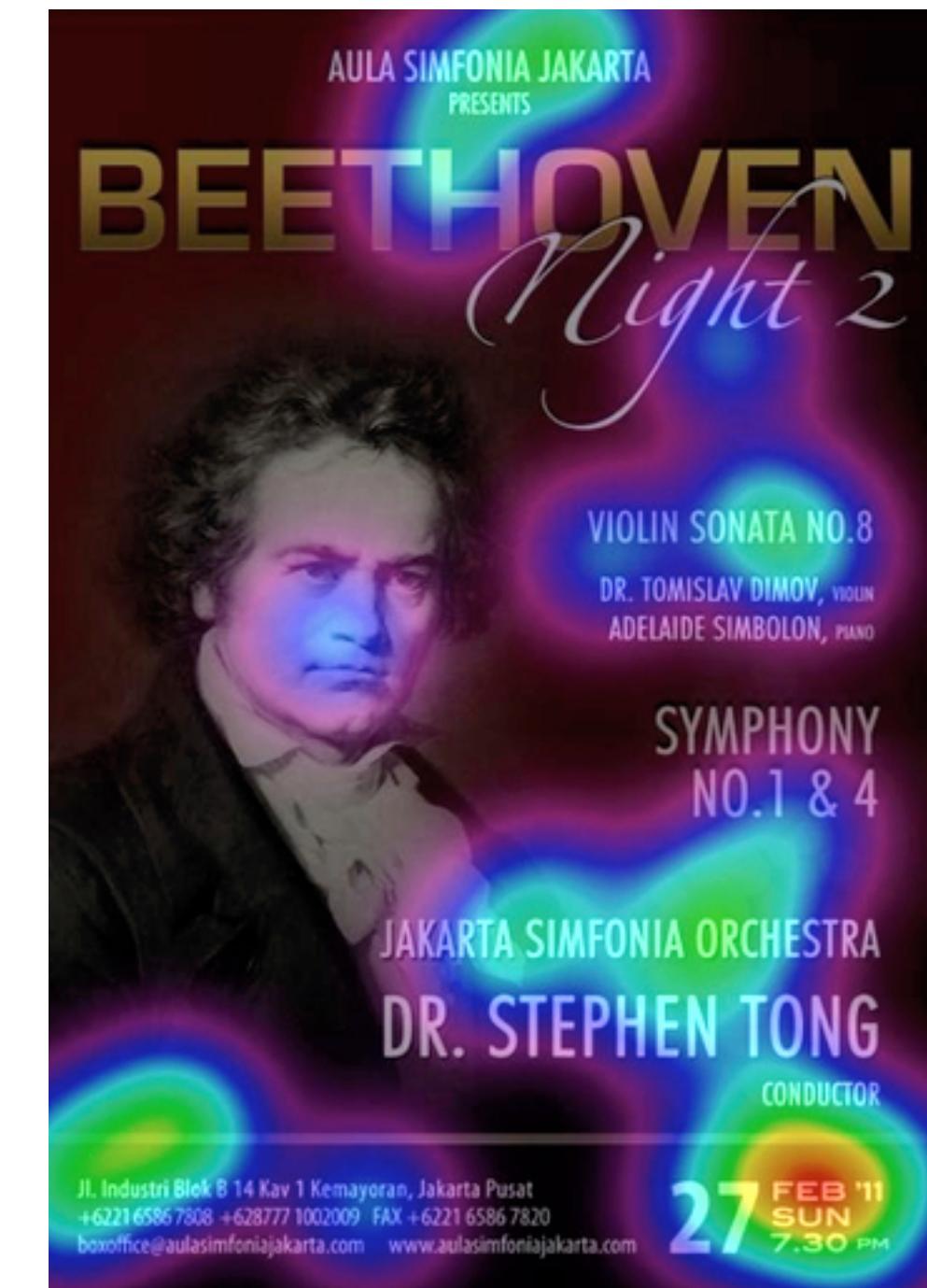
Comparing two versions of a web page to see which one performs better.

Two minor variants of a web page

- Show design A to every even-numbered visitor to web site
- Show design B to every odd number
- Monitor site to see which has higher dwell rate/click-through rate
- Choose better design
- Repeat

Physiological methods

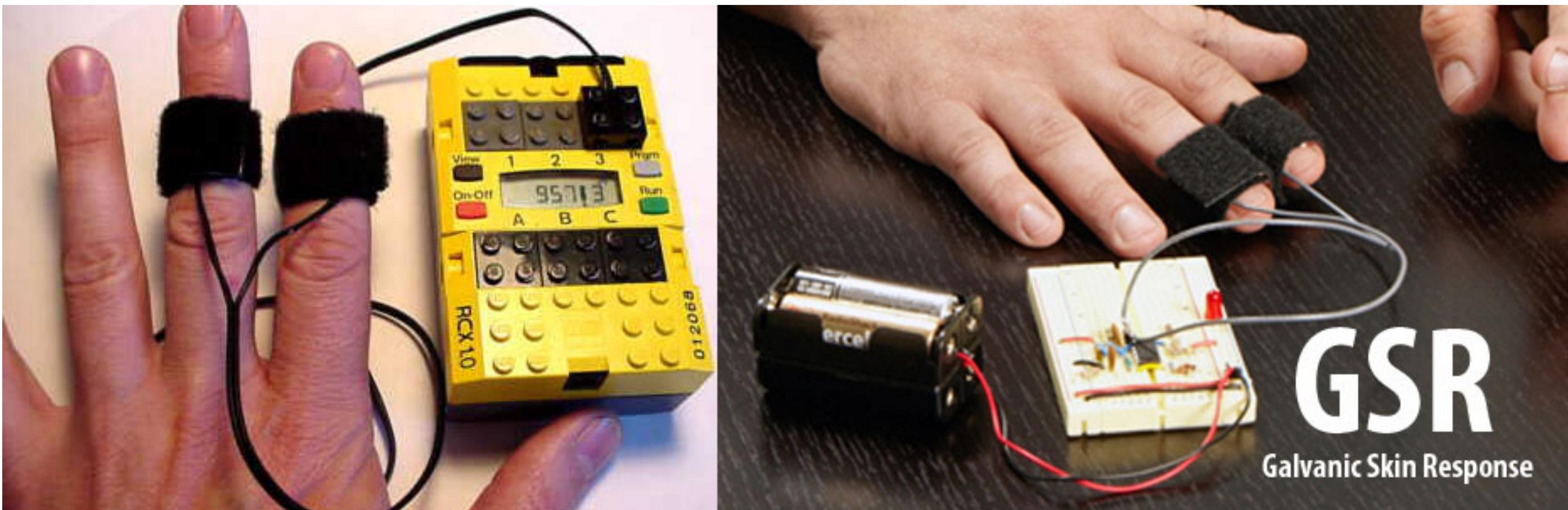
- Eye tracking
- Physiological measurement



Eye tracking

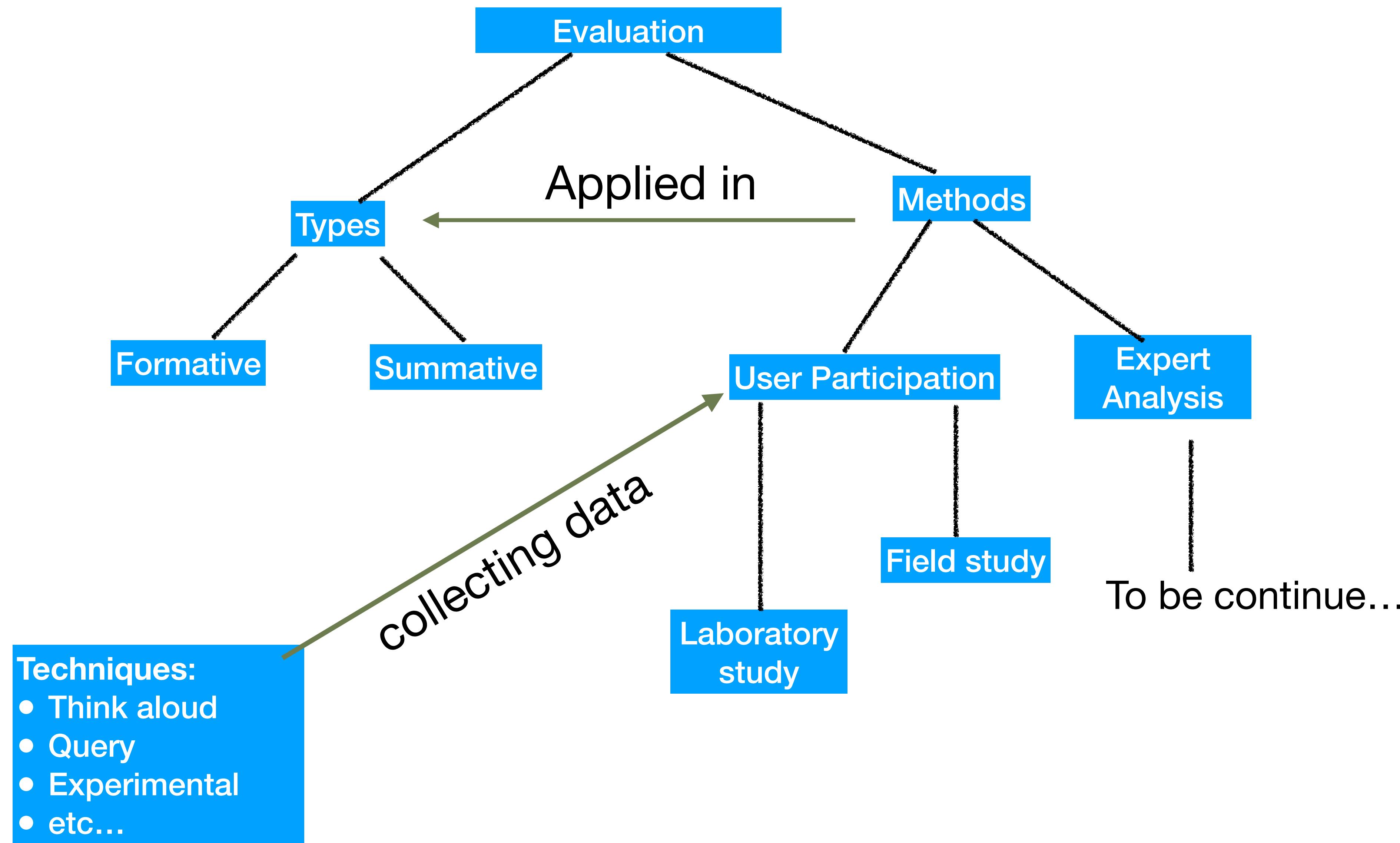
Physiological measurement

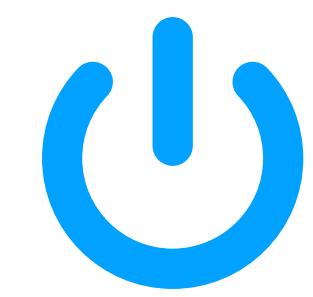
- emotional response linked to physical changes
- these may help determine a user's reaction to an interface
- measurements include:
 - heart activity, including blood pressure, volume and pulse.
 - activity of sweat glands: Galvanic Skin Response (GSR)
 - electrical activity in muscle: electromyogram (EMG)
 - electrical activity in brain: electroencephalogram (EEG)
- some difficulty in interpreting these physiological responses - more research needed



Analysis Data

- Quantitative data analysis
 - Statistical methods – e.g., counting, t-tests, analysis of variance, time-series analysis
- Qualitative data analysis
 - Interpretations, comparative analysis.



 **Thank you**