

Experimental Methods in Computer Science

Departamento de Engenharia Informática, FCTUC, 2023/2024

Experimental Methods in Computer Science (Metodologias Experimentais em Informática)

Henrique Madeira

Master in Informatics Engineering
Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia da Universidade de Coimbra
2023/2024

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering, DEI-FCTUC, 2023/2024

1

1

Experiments with people

Henrique Madeira, DEI-FCTUC, 2018-2023

Experimental Methods in Computer Science, Master in Informatics Engineering, DEI-FCTUC, 2023/2024

2

2

Relevance of experiments with people

- Computer systems and software are developed by people
- Computer systems and software are maintained by people
- Computer systems and software are used by people
- These groups are not the same people
- It is hard to anticipate what other people think and do

Need to experiment to find out

Need to experiment to improve

3

Areas of application

- **Empirical software engineering**
 - How do people design systems?
 - What are good software engineering procedures?
- **Usability testing in HCI**
 - How do people use systems?
 - What are good interface design guidelines?
- **User perception of system use and performance**
 - What do users care about?
- **Security evaluation**
 - Users perception of system security and its impact on usability
- **Software/computer products market assessment**

And social networks,
obviously...

4

Examples of recurring questions

- **Which processes and/or techniques work best?**
 - Example: testing vs. code inspection
 - Example: detailed design vs. agile programming
 - Example: wide menus vs. deep menus
- **Variation between experienced and novice programmers/users**
 - How to make it transparent and easier to learn for newbies?
 - How to make it efficient for experienced users?
 - Can both use the same mechanisms?
- (...)

Experimental techniques

1. Observation and data analysis

- See how users behave on their own (while using computer applications), usually done by monitoring and (big) data analysis

2. Controlled experiments

- See how users perform predefined tasks
- See how user behavior changes when a specific system parameter is changed

3. Interviews and surveys

- Try to understand why users behave the way they do, their preferences, their needs, etc.
- Subjective participant's impression

1 - Observing user behavior: contextual inquiry

Detailed observation of a **small number of people** during their normal work

- What are the real problems?
- What are the real needs?
- Where can you bring real value?
- A deep level of requirements elicitation
 - Users often do not know how to articulate what they need.
 - What they want is not always what they need.
 - What they need may not be computer related... needs observation and analysis to find out.

The system/software/techniques
is the target of the evaluation:
not the users

7

1 - Observing user behavior

Experimental approach: look at user behavior

Goals: understand what users care about in the system/software, their performance and their behavior

- Use logs of system activity, keystroke recording, mouse recording, screen recording, etc.
→ reduce the intrusion as much as possible
- Look at the activity of each user separately → normally the number of participants is small
("observing" massive number of users on the Internet through automatic logging could be done using AI approaches)
- For each user's task, measures users' performance
 - For example, task completion (yes/no), number of mistakes, time on task, etc.
- Correlate the performance with the behavior

8

2 - Controlled experiments with users

Steps in a typical user study:

1. Define the system goals or the module under evaluation – what services/functionalities does it provide?
2. Create a set of tasks that are performed to meet these goals
3. Define measures/observations
 - **Performance** (e.g., task completion, number of mistakes, time on task, etc.)
 - **Subjective opinion** of participants (questionnaires)
4. Get people who are representative of the system users
5. Watch them (or record them) trying to perform the tasks and collect the measurements/observations

Applicable to customers of a web site as well as developers of a new application

9

Experimental aspects

- Measure the performance on the different tasks
 - Average of each task across users
 - Average of each user across tasks
 - Average number of mistakes
 - Etc...
- Record on video (screen capture) for later analysis
 - Screenshots
 - User behavior and expressions
 - Users' mistakes
 - Recording a video is a practical way to measure time on each task
- Collect the subjective experience of the participant
 - Formulate a statement (or a question) about the system
 - Capture testers' opinion using a scale with an **even number of positions** (to avoid the mid point tendency)

10

Experimental aspects

- Measure the time to complete the task
 - Average
 - Average
 - Average
 - Etc...
- Record on the system
 - Screenshots
 - User behaviour
 - Users' mistakes
 - Recording a video is a good way to measure time on each task
- Collect the subjective experience of the participant
 - Formulate a statement (or a question) about the system
 - Capture testers' opinion using a scale with an **even number of positions** (to avoid the mid point tendency)

Example of scale:

- Strongly Disagree
- Disagree
- Slightly Disagree
- Slightly Agree
- Agree
- Strongly Agree

The questions should be provided in the form of statements that can be answered using the proposed scale are. Example of question for a usability test:

- "The system/website is easy to use"

11

Defining the tasks (e.g., for a usability studies)

- Too many tasks... cannot test all
- Make a list of tasks and rate them by importance to the product on a scale of 1 to 6
- Then, rate them by the degree of doubt that the designers (or the owner or informal user's feedback) have about them, again in the scale of 1 to 6
- Multiply the two ratings and sort out the result
- Test the top-ranking tasks: those that are important and most require user input

12

Defining the tasks (e.g., for a usability studies)

- Define the goals first, not the procedure
 - **The goal is to find out what procedure users will use**
- Be specific and very clear about what you want the users do
- Create a reasonable sequence
- Avoid the use of words that appear in the user interface
- Together they should not take too much time
 - Estimate how long it will take you (an expert who knows the system)
 - Multiply by 3 to 10, depending on the tester's profile

Experimental Methods in Computer Science, Master in Informatics Engineering, DEI-FCTUC, 2023/2024

Henrique Madeira, DEI-FCTUC, 2018-2023

13

13

Selecting people (e.g., for a usability studies)

- **Recruiting:** finding some people
 - Based on general demographics and miscellaneous features
 - Age, gender, income level, computer usage level, etc
- **Screening:** finding the right people
 - Filter out those that match the demographics but are probably not useful, for many reasons...
 - Testers should be interested (but not predisposed) in the system, maybe use similar system
 - Testers must be available on the planned test dates
 - Testers must not work in the industry or for owners of competing systems

Experimental Methods in Computer Science, Master in Informatics Engineering, DEI-FCTUC, 2023/2024

Henrique Madeira, DEI-FCTUC, 2018-2023

14

14

Conducting a user session

- Explain that the user is helping to test the system, as opposed to the system testing the user
 - There are no wrong answers
 - If you do not understand, if you have difficulties, it is fine, and in fact our main goal is to know about it
- The user should say out loud what he is trying to do and why
 - Remember, users do not should be ashamed
 - Record the whole process
 - The study organizer is in the background only

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024

Henrique Madeira, DEI-FCTUC, 2018-2023

15

15

Experimental plan

- Better to have multiple small studies than one huge study
- Number of subjects can be as low as 5-6
 - Enough to get feel for results, not necessarily good statistics
- Conduct a pilot session
 - Find out suitability for different user demographics
 - Verify that tasks are reasonable
 - Verify that description of system and tasks are understandable

Experimental Methods in Computer Science, Master in Informatics Engineering , DEI-FCTUC, 2023/2024

Henrique Madeira, DEI-FCTUC, 2018-2023

16

16

3 - Interviews and surveys

- Population being interviewed
 - Sampling bias: you want generally representative users
 - Sample size: the bigger the better, but remember that experiments with people are costly
- Phrasing of the questions
 - Questions should be neutral so as not to affect results
 - Question order is also important
 - Pre-test the questions on a small sample to detect and correct possible problems
- Statistical analysis of results

17

Types of interviews

- Unstructured interview
 - Completely free exchange and information collecting
 - Used as an exploratory tool in initial stages of a study, when the researcher does not know much yet
- Semi-structured interview
 - Basically follow a pre-defined outline of questions
 - Allow user to expand on various topics
 - Also on-line questionnaire where questions depend on previous answers
- Structured interview
 - Filling out a pre-defined questionnaire

18

Questionnaire structure

- Title
- Short introduction – what is this about
- Demographics questions – who is answering
- Start with the easy questions
- Leave sensitive questions to the end
- Be clear and sharp... and pre-test all the questions

19

Questionnaire structure (cont.)

- Questions types:
 - Multiple choice or a scale
 - Numeric (how many times a day do you use Spotify?)
 - Open text (what would you do in the following page?)
- When giving choices, always include "N/A", "other", etc.
- Provide text explanations in addition to scale
- Desirable scale is debatable
 - Number of points should be between 4 and 8
 - Number of points should be even to avoid undecided middle response

20

Ethical considerations

- Ethical considerations are essential. Explicit measures must be taken to prevent problems. But the key ingredient for experiments with people is **trust**.
- Studies often include dealing with confidential information in an organization. This must be taken into account.
- Key ethical factors:
 - Informed consent
 - Review board approval
 - Confidentiality
 - Handling of sensitive results
 - Inducements
 - Feedback