Usability Engineering & User eXperience

Egon L. van den Broek
(met dank aan Christof van Nimwegen)
18 mei 2018

Course book
CHAPTERS 7 + 8
DATA GATHERING &
INTERPRETATION AND PRESENTATION

Extra literature
CHAPTERS 4 + 6
PERFORMANCE METRICS
SELF-REPORTED METRICS
Contents

• Hoofdstuk 7+8 handbook
  – Data gathering
  – Interpretation and presentation

• Hoofdstuk 4+6 Tullis en Albert (2013)
  – Performance metrics
  – Self-reported metrics
Data gathering: Five key issues first

- Interviews
- Observation
- Questionnaires

1. Setting goals
2. Identifying participants
3. Relationship with participants
4. Triangulation
5. Pilot studies
Five key issues: Setting goals

The type of question, e.g.:

1. Did smartphone’s improve our lifes?
2. What menu is better?
3. Is our redesign going the right way?

Once the goal is set you can see which techniques & analyses to be used
1. Once you know the goal, the next step is to know which kind of people you need (population)

2. All of them (saturated sample) is not possible, you need a sample

3. Sampling: choose which participants
   - Probability (statistically robust)
     - Random
     - Stratified
   - Non-probability (statistically less robust)
     - Convenience
Five key issues: Relationship with participants

Informed consent

This is a test of the software. We are not testing you

You can stop at any moment you like

No one will be able to connect your responses or any other information to your identity

Any personal information that could identify you will be removed or changed before files are shared

Your participation in the evaluation is voluntary. You do not have to answer any questions you do not want to and you can stop answering questions at any time just by saying you want to stop.
Five key issues: Triangulation

Look at data from more than one perspective

Collect more than one type of data, e.g. quantitative from experiments and qualitative from interviews.
Pilot studies

• Never omit, but do it in time!!
• Also veterans should always do it
• Test subjects are often costly, don’t waste ‘em
• Most likely some fine tuning suggestions will pop up
• In case of user tests this is your final practice
• Hard/software can have surprises
• It’s not expensive
Data recording

• Notes, audio, video, photographs can be used individually or in combination:
  – Notes plus photographs
  – Audio plus photographs
  – Video
    • 🙁 Intrusive, storage issues
    • 😊 Captures everything

• Different challenges and advantages with each combination
Interviews

- Unstructured - are not directed by a script. Rich but not replicable.

- Structured - are tightly scripted, often like a questionnaire. Replicable but may lack richness.

- Semi-structured - guided by a script but interesting issues can be explored in more depth. Can provide a good balance between richness and replicability.

- Focus groups – a group interview
Interview: questions

Two types:

- **closed questions**: have a predetermined answer format, e.g., ‘yes’ or ‘no’. Closed questions are easier to analyze
- **open questions**: do not have a predetermined format

Avoid:

- Long questions
- Compound sentences - split them into two
- Jargon and language that the interviewee may not understand
- Leading questions that make assumptions e.g., why do you like?
- Unconscious biases e.g., gender stereotypes
- Double negations
- Interruptions

and...

- Let there be silence (e.g., shutdown the smartphones)
- Stay neutral
Interview: running it

- **Introduction** – introduce yourself, explain the goals of the interview, reassure about the ethical issues, ask to record, present the informed consent form.

- **Warm-up** – make first questions easy and non-threatening.

- **Main body** – present questions in a logical order

- **A cool-off period** – include a few easy questions to defuse tension at the end

- **Closure** – thank interviewee, signal the end, e.g. switch recorder off.
Focus Groups

- Often used in marketing, political campaigning, social sciences
- 3-10 people, moderated
- Allows diverse / sensitive issues to be raised
- Individuals develop opinions within a social context (talking)
Questionnaires
Questionnaires

- Questions can be closed or open
- Closed questions are easier to analyze, and may be distributed and analyzed by computer
- Can be administered to large populations
- Disseminated by paper, email and the web
- Sampling can be a problem when the size of a population is unknown as is common online evaluation
Questionnaires: Design

- The impact of a question can be influenced by question order.
- You may need different versions of the questionnaire for different populations.
- Provide clear instructions on how to complete the questionnaire.
- Strike a balance between using white space and keeping the questionnaire compact.
- Avoid very long questionnaires.
- Decide on whether phrases will all be positive, all negative or mixed.
Questionnaires: Question/response format

- ‘Yes’ and ‘No’ check/radio boxes
- Checkboxes that offer many options
- Rating scales
  - Likert scales
  - semantic differential scales
  - 5, 7 or more points (uneven)
- Open-ended responses
Questionnaires: Likert scales

In general, how do you rate the quality of Fictionals chocolate ice cream?

- Poor
- Fair
- Good
- Very Good
- Excellent

<table>
<thead>
<tr>
<th>Poor</th>
<th>Fair</th>
<th>Average</th>
<th>Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. It is the duty of doctors to keep people alive for as long as possible.

- Strongly Agree
- Agree
- Agree somewhat
- Undecided
- Disagree somewhat
- Disagree
- Strongly disagree

Evaluate the following statements.

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I like chocolate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I like peanut butter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) (2) (3) (4) (5)
Questionnaires: Semantic differential

How would you describe Kmart, Walmart, and Target on the following scale:

clean
bright

dirty
dark

Please tell us what you think about Anderson grocery stores.

<table>
<thead>
<tr>
<th>Clean</th>
<th>Friendly employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Describes</td>
<td>No Opinion</td>
</tr>
<tr>
<td>Strongly Describes</td>
<td>Unfriendly employees</td>
</tr>
</tbody>
</table>

Please put a check mark in the space on the line below to show your opinion about the school guidance counselor.

Empathetic  Approachable

Apathetic  Aloof
Questionnaire: Encouraging a good response

• Make sure purpose of study is clear
• Promise anonymity
• Ensure questionnaire is well designed
• Offer a short version for those who do not have time to complete a long questionnaire
• If mailed, include a stamped addressed envelope
• Follow-up with emails, phone calls, letters
• Provide an incentive
• 40% response rate is good, 20% is often acceptable
Advantages of online questionnaires

- Relatively easy and quick to distribute
- Responses are usually received quickly
- No copying and postage costs
- Data can be collected in database for analysis
- Time required for data analysis is reduced
- Errors can be corrected easily
Don’ts! (1)

Don’t write leading questions

- **Bad**: How short was Napoleon?
- **Good**: How would you describe Napoleon’s height?

Stay away from double-barreled questions

- **Bad**: How satisfied or dissatisfied are you with the pay and work benefits of your current job?
- **Good**: How satisfied or dissatisfied are you with the pay of your current job?

People are quite influenceable by the way questions are formulated:

How fast did the red car drive when it
- Hit the grey car
- Crashed against the grey car

Yields different estimations!!
Don’ts! (2)

Be clear by speaking your respondent’s language
- **Bad:** Do you own a tablet PC?
- **Good:** Do you own a tablet PC? (e.g. iPad, Android tablet)

Avoid Ambiguous questions
- **Bad:** Would you be willing to relocate for a better job?
- **Good:** Would you be willing to relocate to another country for a better job? *(depending on what you want to know)*

Do not use absolutes in questions
- **Bad:** Do you always eat breakfast? (Yes/No)
- **Good:** How many days a week do you usually eat breakfast? (Every day/ 5-6 days/ 3-4 days/ 1-2 days/ I usually don’t eat breakfast)
Observation

• Direct observation in the field
  – Degree of participation (insider or outsider)
  – Ethnography

• Direct observation in controlled environments
  – e.g. laboratory, usability lab

• Indirect observation: tracking users’ activities
  – Diaries
  – Interaction logging
  – Video and photographs collected remotely by drones or other equipment
Direct observation in the field
Structuring frameworks to guide field observation

• Three easy-to-remember parts:
  – The person: Who?
  – The place: Where?
  – The thing: What?

• A more detailed framework (Robson, 2014):
  – Space: What is the physical space like and how is it laid out?
  – Actors: What are the names and relevant details of the people involved?
  – Activities: What are the actors doing and why?
  – Objects: What physical objects are present, such as furniture
  – Acts: What are specific individual actions?
  – Events: What you observe, is that part of a special event?
  – Time: What is the sequence of events?
  – Goals: What are the actors trying to accomplish?
  – Feelings: What is the mood of the group and of individuals?
Direct observation in the field: Planning and conducting

• Decide on how involved you will be: passive observer to active participant
• How to gain acceptance
• How to handle sensitive topics (e.g., culture, private spaces)
• How to collect the data:
  – What data to collect
  – What equipment to use
  – When to stop observing
Ethnography is a philosophy with a set of techniques that include participant observation and interviews.

Debate about differences between participant observation and ethnography.

Ethnographers immerse themselves in the culture that they study.

A researcher’s degree of participation can vary along a scale from ‘outside’ to ‘inside’.

Analyzing video and data logs can be time-consuming.

Collections of comments, incidents, and artifacts are made.
Ethnography: Online (2)

• Virtual, Online, Netnography

• Online and offline activity

• Interaction online differs from face-to-face

• Virtual worlds have a persistence that physical worlds do not have

• Ethical considerations and presentation of results are different
Etnography (3): Observations and materials that might be collected (Crabtree, 2007)

- Activity or job descriptions.
- Rules and procedures that govern particular activities.
- Descriptions of activities observed.
- Recordings of the talk taking place between parties.
- Informal interviews with participants explaining the detail of observed activities.
- Diagrams of the physical layout, including the position of artifacts.
- Other information collected when observing activities:
  - Photographs of artifacts (documents, diagrams, forms, computers, etc.)
  - Videos of artifacts.
  - Descriptions of artifacts.
Observation in a controlled environment

• Direct observation
  – Think aloud techniques

• Indirect observation – tracking users’ activities
  – Diaries
  – Interaction logs
  – Web analytics

• Video, audio, photos, notes are used to capture data in both types of observations
Observation in a controlled environment: direct

- Behavior, Utterances, Biofeedback....
Observation in a controlled environment: Direct

Retrospective thinkaloud

Please don't interrupt me while I'm talking to myself.

I noticed that you looked at _____ several times...

What were your impressions of that task?

When you were looking for _____ you didn't seem to notice _____?
Observation in a controlled environment: Indirect

probe  [prōb]

Use probe in a sentence

noun
1. The definition of a probe is an investigation into something or a blunt surgical tool used for exploratory medical testing.

Cultural probes: (Gaver, 1999)
• Focus on someone’s life
• Distribute artefacts
• Follow up interview/debriefing
Observation in a controlled environment: Indirect

Web analytics

• A system of tools and techniques for optimizing web usage by:
  • Measuring,
  • Collecting,
  • Analyzing, and
  • Reporting web data

• Typically focus on the number of web visitors and page views.
Observation in a controlled environment: **Indirect**

Web analytics

50% visitors see variation A

Variation A

23% conversion

50% visitors see variation B

Variation B

11% conversion

Google Analytics

hotjar
Choosing and combining techniques

• Depends on the:
  – Focus of the study
  – Participants involved
  – Nature of the technique(s)
  – Resources available
  – Time available
Quantitative and qualitative data

• Quantitative data – expressed as numbers

• Qualitative data – difficult to measure sensibly as numbers, e.g. count number of words to measure dissatisfaction

• Quantitative analysis – numerical methods to ascertain size, magnitude, amount

• Qualitative analysis – expresses the nature of elements and is represented as themes, patterns, stories
Quantitative and qualitative: After the gathering phase

<table>
<thead>
<tr>
<th>Usual raw data</th>
<th>Example qualitative data</th>
<th>Example quantitative data</th>
<th>Initial processing steps</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interviews</strong></td>
<td>Audio recordings. Interviewer notes. Video recordings</td>
<td>Responses to open questions. Video pictures. Respondent’s opinions</td>
<td>Age, job role, years of experience. Responses to closed questions</td>
</tr>
<tr>
<td><strong>Questionnaires</strong></td>
<td>Written responses. Online database</td>
<td>Responses to open questions. Responses in ‘further comments’ fields. Respondent’s opinions</td>
<td>Age, job role, years of experience. Responses to closed questions</td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td>Observer’s notes. Photographs. Audio and video recordings. Data logs. Think-aloud</td>
<td>Records of behavior. Description of a task as it is undertaken. Copies of informal procedures</td>
<td>Demographics of participants. Time spent on a task. The number of people involved in an activity</td>
</tr>
</tbody>
</table>
Simple quantitative analysis

- **Averages**
  - Mean: add up values and divide by number of data points
  - Median: middle value of data when ranked
  - Mode: figure that appears most often in the data
- **Percentages**
- **Graphical representations give overview of data**
Simple quantitative analysis: be careful with…

- Researcher bias in interpretation
- Think of honest presentation, don’t fool the reader

Not scrutinizing enough (Always study the distribution) \( M=3.5? \)
Simple quantitative analysis: Prepare the data…

- Data entry

- Outliers?
Analysis: Tools to support it

- Spreadsheet – simple to use, basic graphs
- Statistical packages
- Qualitative data analysis tools
  - Categorization and theme-based analysis
  - Quantitative analysis of text-based data
- Nvivo and Atlas.ti support qualitative data analysis
- CAQDAS Networking Project, based at the University of Surrey (http://caqdas.soc.surrey.ac.uk/)
Look!

Figure 8.7 Interaction profiles of players in the starport

Simple qualitative analysis

- Recurring patterns or themes
  - Emergent from data, dependent on observation framework if used
Simple qualitative analysis: Video
Simple qualitative analysis

• Categorizing data
  – Categorization scheme may be emergent or pre-specified

00:01:15 user goes to help screen
“How do I move to the previous screen?”
NAVIGATION - PROBLEM

“I have diabetes you know and I’m hoping this can help me. So here I go, I’m clicking on this screen about diabetes information”

00:01:23 subject goes to diabetes help screen and clicks on help button

“How now what? It looks like everything has stopped”
LACK OF INDICATION OF SYSTEM STATUS – PROBLEM

– Het staat een beetje verborgen in een ingeklapte menu. Ik had het eerder terug verwacht op een pagina.

– Ik moest er even naar zoeken doordat ik het hamburgermenu niet direct zag en op zoek was naar een overzicht/userinfo knop.
– Ik vond het nog een beetje onduidelijk wat het verschil was tussen performance en progress.
Simple qualitative analysis

- Looking for critical incidents (or: anomalies)
  - Helps to focus in on key events

Advantages
- Focuses on important issues e.g. safety critical events may bring major benefits
- Useful for identifying rare events that might not be picked up by other methods

Disadvantages
- Routine incidents may not be reported
- Poor as a tool for general task analysis
- Critical incidents often rely on memory, incidents may be distorted or even forgotten if they are collected long after an event.
Presentation: The right choice....?
Presentation: the right choice....?
Chapter 4: performance metrics
How sacred is it 😊?

There are 3 kinds of lies: 

Lies, damn lies, and statistics

HOW TO LIE WITH STATISTICS
Darrell Huff
Illustrated by Irving Geis
Performance metrics

- A bit farther from UX, a bit closer to UE
- Adequate size needed
- They tell what, not why

1. Task success
2. Time on task
3. Errors
4. Efficiency
5. Learnability
Performance metrics: Confidence intervals

95% confidence interval: 11.2 - 18.7

90% confidence interval: 11.9 - 18.1
Performance metrics: Task success

- Binary: yes or no, 0 or 1

- Levels of success
  - Vary e.g. user’s experience in a task
  - Vary level of optimality

- Issues
  - How to define success
  - Strikes and out?
  - “Call” it at a certain point?
Performance metrics: Time on task

• Informs on efficiency
• Often: the faster the better (not always)
• What about…engagement?
Performance metrics: Errors

- When to measure them?
- What is an error
- Collecting and measuring them
- Analyzing – presenting errors
- Issues to consider
Performance metrics: Efficiency

- Steps: path measures
Casus: Performance metrics - Destination feedback

- Destination feedback
- Slots in schedule (that are possible AND available) turn green on click
- Not the best slots, simply the one that are possible

NO destination feedback
- User must look at the limitations and think for him/herself and plan ahead
Casus: Time on task

No difference in average time needed

Time before first move (s)

- NO FEEDBACK: Mean = 19.8
- DESTINATION FEEDBACK: Mean = 15.3

NO FEEDBACK waits longer than DESTINATION FEEDBACK \( (p < .05) \)

Time between moves (s)

- NO FEEDBACK: Mean = 4.8
- DESTINATION FEEDBACK: Mean = 3.9

NO FEEDBACK takes longer between moves than DESTINATION FEEDBACK \( (p < .05) \)
Performance metrics: Learnability

• More “self-service” for all of us (travel, banking products, tax, e-government, e-health, insurances, ……)
• In casus?
Performance metrics: Learnability in Casus

Vertical: Superfluous moves (from ideal path)
Chapter 6: Self reported measures

- UX: Shift from more quantitative to more qualitative
- Accent on performance measures is a human factors legacy
- Delight, joy, trust, fun etc. receive more attention
Self reported measures

• Subjective data, preference data?
• User’s perceptions
• When: during, post-task, post-study
• Social desirability bias (e.g. phone more positive than form)

What?
• attitudes
• Beliefs
• Opinions
• …. 
Likert or Semantic Differential?

Likert scale
+ is easier to answer (seems)
- It has a negative polar and positive polar

Semantic Differential
+ without prejudgment (the word, e.g. I *enjoy* using this system)
- The divisions can be a problem
Post session ratings

- After the complete range of interactions
  - Lab = free again
  - Overall barometer

- Aggregated individual task ratings
- System Usability Scale (SUS)
  - <50: not acceptable
  - 50-70: marginal
  - >70 Acceptable
**System Usability Scale**

**Post session ratings: SUS**

1. I think that I would like to use this system frequently
   - Strongly disagree: 1, 2, 3, 4, 5
2. I found the system unnecessarily complex
   - Strongly disagree: 1, 2, 3, 4, 5
3. I thought the system was easy to use
   - Strongly disagree: 1, 2, 3, 4, 5
4. I think that I would need the support of a technical person to be able to use this system
   - Strongly disagree: 1, 2, 3, 4, 5
5. I found the various functions in this system were well integrated
   - Strongly disagree: 1, 2, 3, 4, 5
6. I thought there was too much emphasis on unnecessary functions
   - Strongly disagree: 1, 2, 3, 4, 5
Post session ratings: SUS

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
Post session ratings

- Computer system Usability Questionnaire (QSUQ)
  - 19 statements, 5-point Likert
- Questionnaire for User Interface Satisfaction (QUIS)
  - 27 scales (10-point)
- Usefulness, Satisfaction and Ease of use Questionnaire (USE)
  - 30 (7-point) Likert rating scales, in 4 categories
- Product Reaction Cards
  - 118 cards, choose the ones describing your experience, then explain top 5
Post session ratings: Net Promoter Score

- Net promoter score (NPS)
  - “how likely is it that you would promote, 1-item scale

- Sauro (2010): SUS and NPS correlate $r=0.61$, $p<0.001$
- SUS scores: Promotors: 82, Detractors: 67