HCI evaluation

part two

Stuart Gray, Pete Bennett, Simon Lock, Thomas Bale, Harry Field

Summary

In **HCI (part 1)** we covered:

- Questionnaires
- Observations
- Interviews
- Focus groups
- Ethnography

Today in **HCI (part 2)** we will be looking at:

- What to do with the data from HCI part one?
- How do you evaluate a working system?

Interaction Design

Again... there's much more detail in the book. We'll be giving a high level overview here, but please refer to the book for in-depth discussion of methods.

Chapter 8 - Data Gathering

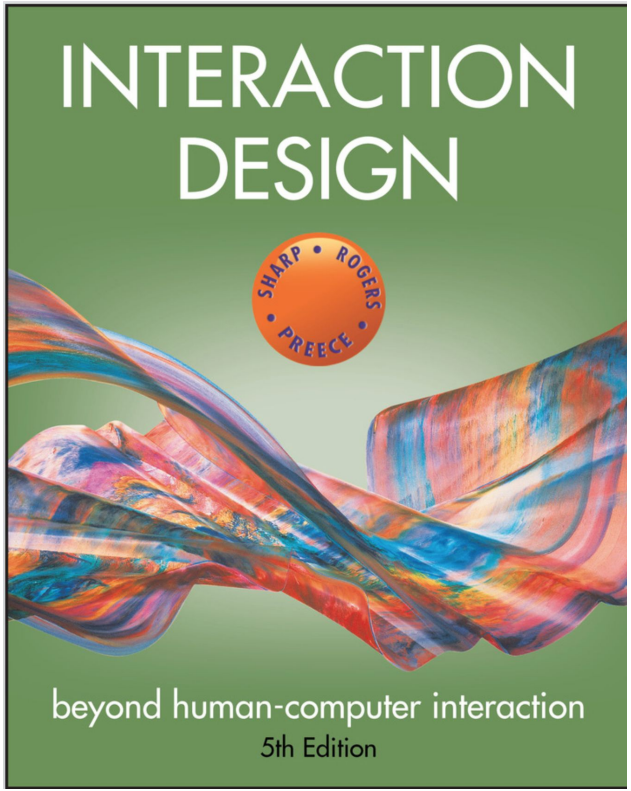
Chapter 9 - Data Analysis, Interpretation, and Presentation

Chapter 10 - Data at Scale

Chapter 14 - Introducing Evaluation

Chapter 15 - Evaluation Studies: from Controlled to Natural Settings

Chapter 16 - Inspections, analytics and models



Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. Interaction Design: Beyond Human - Computer Interaction(3rd. ed.). Wiley Publishing.

Types of evaluation

Controlled settings directly involving users (examples are usability labs and research labs): Users' activities are controlled to test hypotheses and measure or observe certain behaviors. The main methods are usability testing and experiments.

Natural settings involving users (examples are online communities and products that are used in public places): There is little or no control of users' activities to determine how the product would be used in the real world. The main method used is field studies (for example in-the-wild studies).

***Any settings not directly involving users*:** Consultants and researchers critique, predict, and model aspects of the interface to identify the most obvious usability problems. The range of methods includes inspections, heuristics, walk-throughs, models, and analytics.

Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. Interaction Design: Beyond Human - Computer Interaction(3rd. ed.). Wiley Publishing.

Controlled Settings Involving Users

Controlled settings directly involving users (examples are usability labs and research labs): Users' activities are controlled to test hypotheses and measure or observe certain behaviors. The main methods are usability testing and experiments.

- **Goal of controlled studies**
  - To test whether the product being developed (or a specific element of it) is usable by the intended user without the presence of external variables
  - To understand whether users are satisfied with their experience
  - To compare different products or different product versions
  - To pose and test hypotheses (formal predictions we make) about our product and/or users

Controlled Settings Involving Users

- **Methods**
  - Usability testing, any of the methods already described but in a more controlled setting - i.e. observations, performance measures, interviews, A/B testing
  - If you are taking an experimental approach (formally testing one or more hypotheses and common within A/B testing), you may need to consider the following elements of study design to ensure statistical validity - presence and activities of control groups, randomisation of participants, condition ordering, sample sizes, independent and dependent variables, performance benchmarking

- **Some examples of performance measures**
  - Time to complete a task
  - Number and type of errors per task
  - Number of navigations to online help or manuals

- **Notes** - You are unlikely to be doing formal, tightly controlled evaluation during this project, but you may have undertaken similar activities in a more informal manner (‘opportunistic evaluations’)

Controlled Settings Involving Users

- **Positives**
  - Removes the effects of many outside influences on users
  - More objective for researchers and users - elimination of many biases
  - Easier to analyse as data is uniform in appearance
  - Can allow for more statistically robust results

- **Negatives**
  - Can be costly / time consuming - e.g. may need to pay users, Get ethical approval / consent, develop session protocols
  - Decontextualised from real world use
  - researcher/observer effects - users are in unfamiliar territory

Natural Settings Involving Users

Natural settings involving users (examples are online communities and products that are used in public places): There is little or no control of users' activities to determine how the product would be used in the real world. The main method used is field studies (for example in-the-wild studies).

- **Goal of field studies**
  - Help identify opportunities for new technology
  - Establish the requirements for a new design
  - Facilitate the introduction of technology or inform deployment of existing technology in new contexts
  - To understand how your design works 'in the wild'

Natural Settings Involving Users

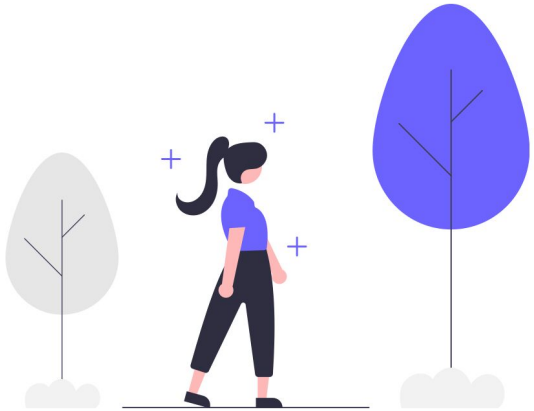
- **Methods**
  - Again, many of the same ones already mentioned, greater emphasis on observation and ethnography.
  - These field studies are often (but not always) undertaken ‘longitudinally’ - run over an extended time period rather than a single evaluation session.
  - Thus, data may go further than acute data measures, to consider patterns of use / performance over time.
  - Users can take an active role in data collection - e.g. user diaries - where users help to track and record their own use of your product and / or their usage context.

- **Notes**
  - This will likely be a good choice for evaluating your completed web app.
  - E.g. with users evaluating your web app in home contexts over a pre-defined time period.

Natural Settings Involving Users

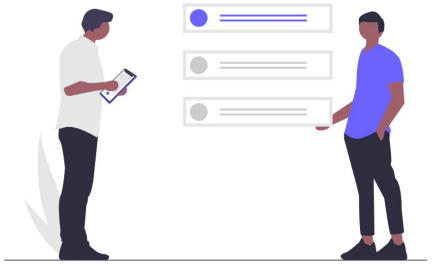
- **Positives**
  - More representative of end usage context - allows you to study logistical, social, and environmental influences upon your design.
  - More comfortable setting for participants, with more realistic user motivations - may counter researcher effects.

- **Negatives**
  - Giving up control makes it difficult to anticipate events.
  - As usage context varies across different participants, it adds additional variables you cannot control for and can make it difficult to provide solid reasons for phenomena.



Any settings not directly involving users

- **Methods**
  - Heuristics - applies knowledge of typical users guided by rules of thumb or through consultations with expert user researchers to evaluate a design. Nielsen (1994) heuristics are a commonly used set of generalisable design heuristics; similarly Budd (2007) has a set that focus on web design.
  - Cognitive walkthroughs - emulating a user's problem-solving process at each step in the human-computer dialogue and checking to see how users progress from step to step in these interactions.
  - Analytics - a technique for logging and analyzing data about user behaviours while using our product and / or our product visibility on the web.
  - Models (of behaviour) - leaning on the underlying empirical research models of human behaviour using similar technologies / designs, can be used to predict or estimate evaluative outcomes. See Fitts' law (1954) as an example.

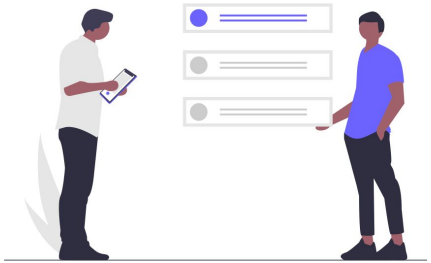


Any settings not directly involving users

- **Positives**
  - When undertaken remotely, analytics can help bring the lab to naturalistic environments.
  - Analytics can automate the collection of many performance measures that may be observation intensive.
  - Models and heuristics can help in predicting user behaviour and when compared against primary data, can uncover important research questions / hypotheses.

- **Negatives**
  - Heuristics commonly make many assumptions about products and people.
  - Heuristics may not be suited to novel technologies - there is a risk of new technology outpacing heuristics.

- **Notes**
  - This could also work well for your teams in your own projects.



Choosing and Combining Methods

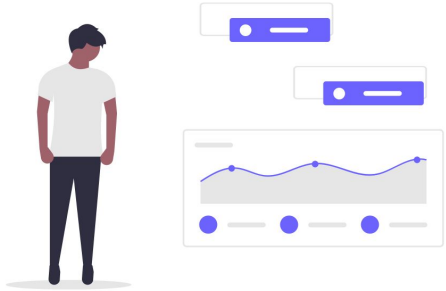
- The methods and settings described are often used to compliment each other and both controlled, natural, and the additional approaches to evaluation may be used at different stages of a project.

- Commonly, for example, while one may conduct controlled evaluation of their product during development, it may not be of a more exploratory nature at the early stages, before becoming more formal with a greater number of controls as the team becomes more confident in formulating and testing hypotheses.

- Similarly, ‘in the wild’ field studies may sit alongside controlled evaluative environments. These are also frequently conducted after a product has been validated within controlled environments.

Analysing Your Data

- Is your data quantitative, qualitative or do you have both?
- Quantitative methods generate data that are represented by measured quantities and is thus numerical. It is best suited when you want to statistically define objectively measurable ‘facts’ about your design or users.
- Qualitative methods generate data that are represented by observed or reported words and/or behaviours. It is best suited when you want to understand and represent the rich opinions and perspectives of your users.
- Mixed methods generate both quantitative and qualitative data, often using one to inform the other (‘triangulation’).



Analysing Your Qualitative Data

- Qualitative data may include - transcribed interviews, field observation notes, open-ended survey responses, user-diaries.

- There are several ways in which such data may be analysed, with differing levels of rigour. However, thematic analysis is a popular and flexible means of applying structure and a prescribed process for your analysis.

- In short, this approach involves:
  - Gathering together your data from all available sources
  - Creating category labels (‘codes’) that describe emerging and/or expected patterns / trend
  - Apportioning evidence (e.g. quotes from users, research observations excerpts) underneath each category label
  - Restructuring codes into a hierarchy tree and grouping related items together to generate themes

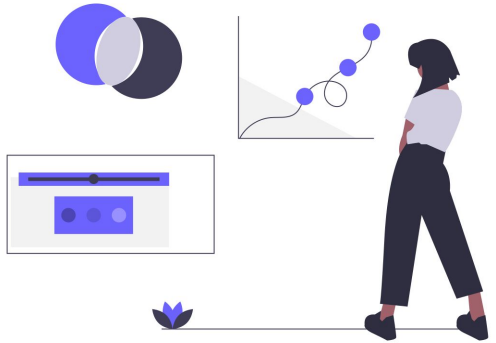
- For a full walk-through find the bonus video on thematic analysis on Teams. Also consider looking into Affinity Diagrams for an alternative means of representing your data and identifying trends.

Analysing Your Quantitative Data

- There are many different statistical tests that one might use to analyse quantitative data that has been collected under controlled conditions. Which one to use will depend on the specifics of your experiment. Nevertheless, in Human-Computer Interaction, you are likely to use both descriptive and inferential statistics.

- Descriptive statistics are used to describe the basic features of the your data and summarise the sample (the number of users participating) and the measures.

- Inferential statistics allow you to make predictions (“inferences”) from described data. **Note**, although incredibly powerful, we will steer clear of inferential statistics within the context of this unit. This is a more advanced skills that is a little beyond our scope.



Analysing Your Quantitative Data

- Common descriptive statistics that may be useful to describe your data:
  - **Sums and totals** - e.g. the number of users who involved in your evaluation; the total number of errors made by all the users
  - **The distribution** - a summary of the frequency of individual values or ranges of values for a variable
    - Frequency distribution tables or graphs are often used to visually show the range of values of a single variable
    - Percentages are also used to show distributions
  - **Central tendencies** of a distribution show where the centre lies in our data using means, medians, and modes
  - **Dispersion** tells us more than just the range of highest and lowest values, calculating standard deviations allows us to understand how much variance there is from the mean - this is important to understand if you plan to undertake any inferential statistics
  - **Correlations** are used to measure how strong a relationship is between two variables. E.g. a correlation between age and self-reported level of enjoyment of your web app.

References

- Nielsen, J., 2005. Ten usability heuristics. https://pdfs.semanticscholar.org/5f03/b251093aee730ab9772db2e1a8a7eb8522cb.pdf
- Budd A. Heuristics for Modern Web Application Development, 2007. URL http://www.andybudd.com/archives/2007/01/heuristics_for_modern_web_application_development. 2007
- Krug, S., 2000. Don't make me think!: a common sense approach to Web usability. Pearson Education India.
- Field, A. and Hole, G., 2002. How to design and report experiments. Sage. Vancouver
- Rogers, Y., Sharp, H. and Preece, J., 2011. Interaction design: beyond human-computer interaction. John Wiley & Sons.