

Sun 5 Apr

Lecture 11: Evaluation (part 1)

Overview

- Introduction
- Why, What, Where & When of evaluation
- Types of Evaluation
- Evaluation case study
- What did we learn from the case study?

Based on material from:

Sharp, H., Preece, J. and Rogers, Y. (2019). *Interaction Design Beyond Human-Computer Interaction*. 5th Ed. Sussex: John Wiley & Sons Ltd.

Learning Outcomes

- Explain the key concepts and terms used in evaluation
- Introduce range of different types of evaluation methods.
- Show how different evaluation methods are used for different purposes at different stages of the design process and in different contexts of use.
- Show how evaluators mixed and modified to meet the demands of evaluating novel systems.
- Discuss some of the practical challenges of doing evaluation.

1) Introduction

- Evaluation is integral to the design process
- Evaluators collect information about users' or potential users' experiences when interacting with a prototype, a computer system, an application or a design artifact
- They do this because to improve design *interviews were eye-openers*
- Evaluation focuses on both the usability of the system (e.g. how easy to learn and use) and the user experience when interacting with the system (e.g. how satisfying, enjoyable or motivating the interaction is)

1) Introduction (cont)

“User experience (UX) is an approach to product development that incorporates direct user feedback throughout the development cycle (human-centered design) in order to reduce costs and create products and tools that meet user needs and have a high level of usability (are easy to use)”

User experience Professionals Association (UXPA)

<http://uxpa.org/resources/about-ux>

2. Why, what, where & when to evaluate

Iterative design & evaluation is a continuous process that examines:

- Why: to check users' requirements and that users can use the product and they like it.
- What: a conceptual model, early prototypes of a new system and later, more complete prototypes.
- Where: in natural and laboratory settings.
- When: throughout design; finished products can be evaluated to collect information to inform new products.

3. Types of evaluation

- Controlled settings that directly involve users (eg. usability and research labs).
- Natural settings involving users (eg online communities and products that are used in public places). Often there is little or no control over what users do, especially in in-the-wild settings.
- Any setting that doesn't directly involve users (eg consultants and researchers critique the prototypes, and may predict and model how successful they will be when used by users.

Usability Lab



Living Lab

- People's use of technology in their everyday lives can be evaluated in living labs.
- Such evaluations are too difficult to do in a usability lab.
- Eg: the Aware Home was embedded with a complex network of sensors and audio/video recording devices (Abowd et al., 2000).
- More recent examples include whole blocks and cities that house hundreds of people, eg Verma et al's, 2017 research in Switzerland.
- Many citizen science projects can also be thought of as living labs eg iNaturalist.com
- These examples illustrate how the concept of a lab is changing to include other spaces where people's use of technology can be studied in realistic environments.

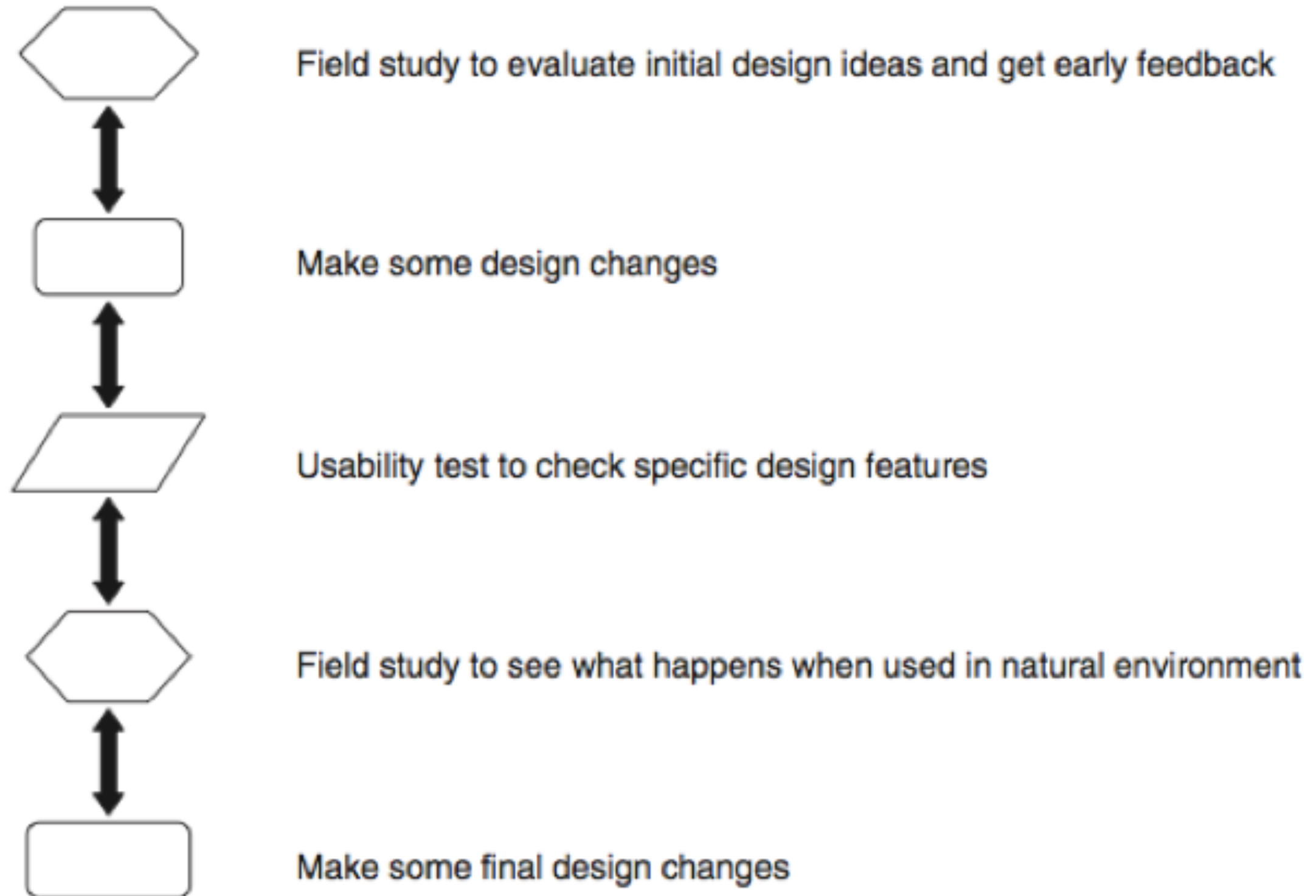
Evaluation Case Studies

natural cause
Doesn't indicate location

- ① • A classic experimental investigation into the *natural* physiological responses of players of a computer game.
- ② • An ethnographic study of visitors at the the Royal Highland show in which participants are directed and tracked using a cell phone app. ~~*Natural*~~ *Controlled since at location*
- ③ • Crowdsourcing in which the opinions and reactions of volunteers (ie the crowd) inform technology evaluation.

Controlled

3.1 Usability testing & field studies can compliment



4. Case study: Airline booking system

Testing Cycle

1. Test Instruction



2. Testing with Users



3. Research & Analysis

4.1 Task Instruction

Task Oriented

- E.g. XXX booking system
- Typical instruction such as:
 - Get a ticket to London on certain date

Non-leading questions

- Non-leading: Try to buy plane tickets for your family vacation
- Leading: Press the red button and then click the 1st top right button with an airplane icon..

4.1 Task Instruction (cont)

- **Sample test instruction**

- Tasks:

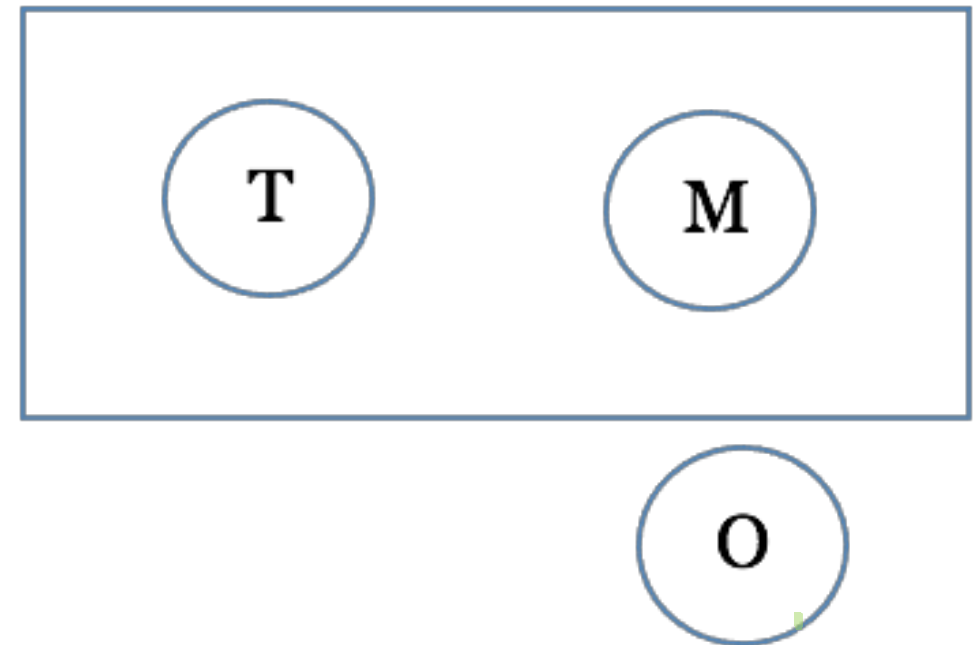
- What impression does this website give you compared to other airline websites? *Non*
- You want to fly to Chicago this January for 5 days. Find suitable flight times (return tickets) for you and your family *leading*
- Decide on suitable flight times and try to buy the tickets. Tell us if anything that will affect you from completing the purchase *Non*
- How many luggage can you bring with you on this flight?

4.2 Testing with users

In person testing

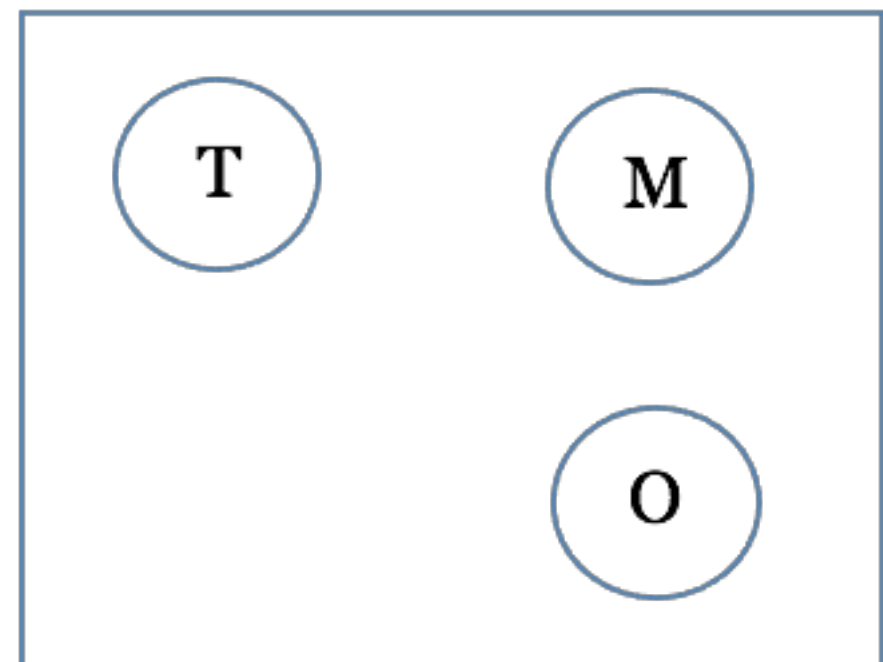
- 1) Tester
- 2) Moderator
- 3) Observer

Setting 1



OR

Setting 2



4.2 Testing with users (cont)

Tester → user

- 1) Think Aloud
- 2) Be >100% honest (likes/dislikes; difficult/easy)
- 3) It is not IQ test
- 4) Give suggestion

4.2 Testing with users (cont)

Moderator

- 1) Tour guide
- 2) Interrupt only when necessary
- 3) Dont answer questions for tester
- 4) Ask questions starting with “What” or “How” if tester becomes quiet

4.2 Testing with users (cont)

Observer

1) Watch

2) Listen

3) Take notes

4.2 Testing with users (cont)

Remote UX testing by crowd sourcing

- Conduct testing using crowd sourced testers so that clients can watch & listen
- People involve:
 - Observer
 - Tester



5. Research & Analysis

How to analyze results?

1. Quantitative - expressed as numbers
2. Qualitative - an inquiry into the reasoning behind human behavior

(as discussed in the previous lecture)

5. Research & Analysis (cont)

Common practice for data analysis process:

1. Transcribe the evaluation sessions (verbal and actions both from the tester & the moderator)
2. Analyze the transcript based on the objective of the evaluation (e.g usability, user behavior etc)
3. Highlight other issues observed from the sessions
4. Report findings

Evaluation methods

Type of evaluation (col)

Method	Controlled settings	Natural settings	Without users
Observing	X	X	
Asking users	X	X	
Asking experts		X	X
Testing	X		
Modeling			X

Testing
w/
users
(row)

only
work
in
controlled

observe
can't do
w/out
users

Participants' rights and getting their consent

- Participants need to be told why the evaluation is being done, what they will be asked to do and their rights.
- Informed consent forms provide this information and act as a contract between participants and researchers.
- The design of the informed consent form, the evaluation process, data analysis and data storage methods are typically approved by a high authority, eg. Institutional Review Board.

Things to consider when interpreting data

- Reliability: does the method produce the same results on separate occasions?
- Validity: does the method measure what it is intended to measure?
- Ecological validity: does the environment of the evaluation distort the results?
- Biases: Are there biases that distort the results?
- Scope: How generalizable are the results?

Key points

- Evaluation & design are closely integrated in user-centered design.
- Some of the same techniques are used in evaluation as for establishing requirements but they are used differently (e.g. observation interviews & questionnaires).
- Three types of evaluation: laboratory based with users, in the field with users, studies that do not involve users
- The main methods are: observing, asking users, asking experts, user testing, inspection, and modeling users' task performance, analytics.
- Dealing with constraints is an important skill for evaluators to develop.

Summary

- Test instruction: goal oriented
- Usability testing can be done easily
- User's action speaks louder than words
- Testing can be done in any stages
- You are not your user
- Testing with one user better than none

Link

[Japan's technology leads the way in caring for the elderly - YouTube](#)