

Topic 8: Introducing Evaluation (Part 1)

SECV2113 Human-Computer Interaction

Faculty of Computing

Universiti Teknologi Malaysia



01 OVERVIEW

02 TYPES OF EVALUATION

03 EVALUATION CASE STUDIES

04 OTHER ISSUES TO CONSIDER

OVERVIEW



Why, What, Where, and When to evaluate

Iterative design and evaluation is a **continuous process**:

- **Why:** To check that all aspects of the **user experience** are taken into account in design
- **What:** A **conceptual model**, early and subsequent prototypes of a new system, more complete prototypes, and to compare a design with competitors' products
- **Where:** In natural, in-the-wild, in living lab and lab settings
- **When:** Throughout design; finished products can be evaluated **to collect information** to inform new products and **for comparisons** with other products

TYPES OF EVALUATIO N

....

Evaluation Types

- **Controlled settings** that directly involve users (for example, usability and research labs)
- **Natural settings** involving users (for instance, 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** (for example, consultants and researchers **critique the prototypes**, and may predict and model how successful they will be when used by users)
- **Remote evaluation** can be done with studies in **all three categories above**.
- Remote evaluation was practiced since the 1990s but it became important recently for coping with **the impact of the Covid-19 pandemic**
- **Different methods** may also be used together **to provide different perspectives**. For example, testing in a lab maybe used to compare different parts of a design, such as two different menu structures. Then the design may be tested in a natural setting to see whether people like using it in their homes or in other natural settings

Living Labs

- People's use of technology in their everyday lives can be evaluated in living labs. For example, evaluators can see **how people use technology over a period of time, and how it fits in/or not with their routines**
- Such evaluations are **too difficult to do in a usability lab**
- An early example was the Aware Home that was embedded with a complex network of sensors and audio/video recording devices (Abowd et al., 2000)

Living Labs (cont.)

- More recent examples include whole blocks and cities that house hundreds of people, for example, Verma et al., research in Switzerland (2017), and the huge PEARL facility in the UK
- Many citizen science projects can also be thought of as living labs, for instance, [iNaturalist.org](https://www.inaturalist.org)
- 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

EVALUATIO N CASE STUDIES

....

Contrasting Case Studies

- A study to test how well an automated vehicle supports drivers' behavior as they transition control to and from between the automated vehicle and the driver.
- An **ethnographic study** of visitors at the the Royal Highland show in which participants are directed and tracked using a mobile phone app
- **Crowdsourcing** in which the opinions and reactions of volunteers (for example, from the crowd) inform technology evaluation

DeepTake and Automated Vehicles

- **Large amounts of data** were collected about drivers' driving and non-driving behavior and analyzed using machine learning techniques
- A **user study** was conducted to examine driver takeover behavior – ie, when a driver took control of the vehicle and when they left control to the automated system
- 11 females and 9 males aged 18-30 years **performed tasks of varying difficulty**
- Each task took about an hour and was followed up by a questionnaire. Eye tracking was used to track where the driver was looking and the time and processes to complete each task was recorded

The user study set-up



Source: Pakdamanian et al., 2021. CHI'2021 <https://dl.acm.org/doi/pdf/10.1145/3411764.3445563>

Ethnobot: Gathering Ethnographic data at the Royal Highland Show

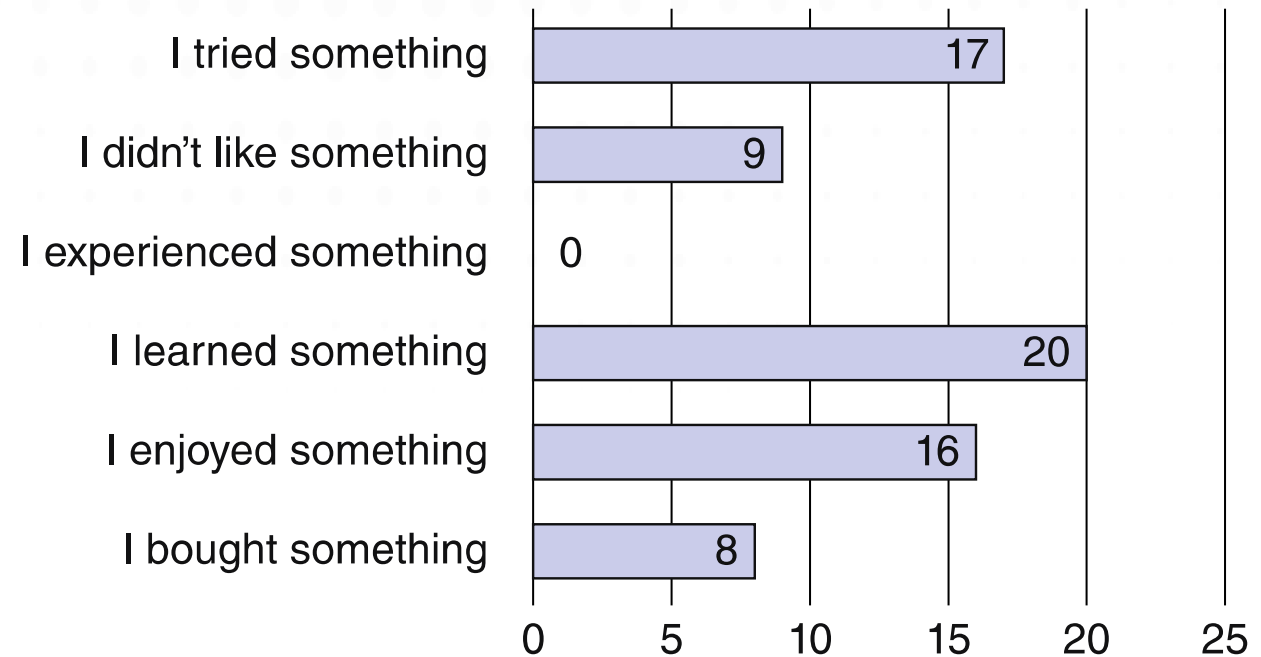
- The Ethnobot directed Billy to a particular place (Aberdeenshire Village)
- Next, Ethnobot asks “...what’s going on?”
- The screen shows five of the experience buttons from which Billy needs to select a response



Source: Tallyn et al. (2018) Reproduced with permission of [ACM Publications](#).

Experience responses to questions submitted in Ethnobot

Number of prewritten experience responses submitted by participants to the pre-established questions that Ethnobot asked them about their experiences



Source: Tallyn et al. (2018) Reproduced with permission of [ACM Publications](#).

Mechanical Turk Test - Amazon



What did we learn from the case studies?

- How to observe users in a lab simulation and in natural settings
- How evaluators use different levels of control in the lab and in natural settings and in crowdsourcing evaluation studies
- Use of different evaluation methods
- How to develop different data collection and analysis techniques to evaluate user experience goals such as engagement
- The ability to run experiments on the Internet that are quick and inexpensive using crowdsourcing
- How a large number of participants can be involved in crowdsourcing using systems such as Mechanical Turk

Evaluation Methods

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

The Language of Evaluation

- Analytics
- Analytical evaluation
- Biases
- Controlled experiment
- Crowdsourcing
- Distributed evaluation
- Ecological validity
- Expert review
- Field study
- Formative evaluation
- Heuristic evaluation
- Informed consent
- In the wild evaluation
- Living lab
- Pain points
- Predictive evaluation
- Reliability
- Research in the wild studies
- Scope
- Summative evaluation
- Usability lab
- User studies
- Usability testing
- Users or participants
- Validity

OTHER ISSUES TO CONSIDER

....

Participants' Rights and Getting Consent

- Participants need to be told **why the evaluation is being done**, what they will be asked to do and informed about 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 higher authority**.

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?

Summary

- **Evaluation and design** are very closely integrated
- Some of the same data gathering methods are used in evaluation as for **establishing requirements and identifying users' needs**, for example, observation, interviews, and questionnaires
- Evaluations can be done in **controlled settings** such as lab, less controlled field settings, and remotely or without users being involved
- **Usability testing and experiments** enable the evaluator to have a high level of control over what gets tested, whereas evaluators typically impose little or no control on participants in field studies, particularly in the wild studies
- **Different methods can be combined** to get different perspectives
- Participants need to be made **aware of their rights**
- It is important **not to over-generalize findings** from an evaluation