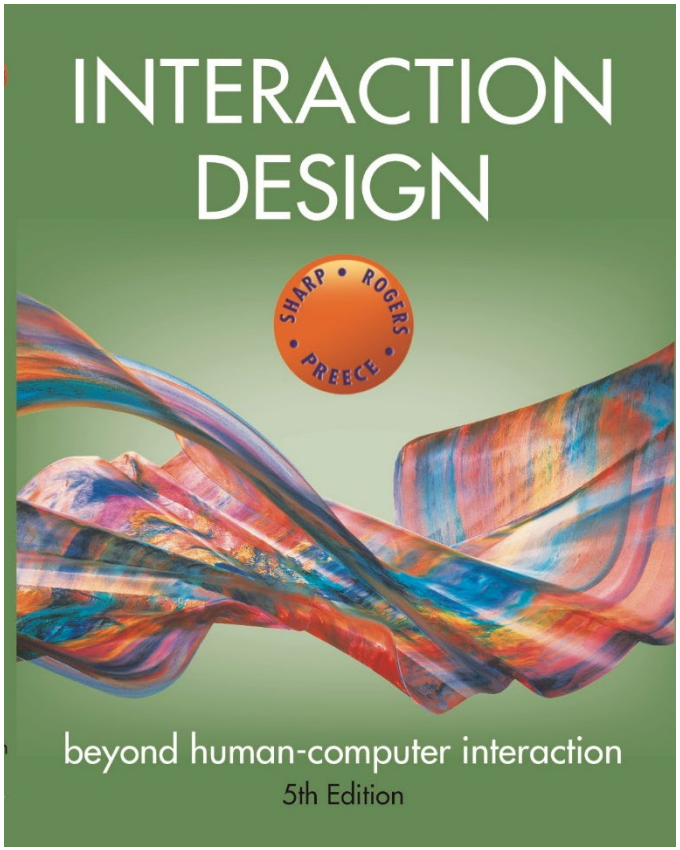


Chapter 14
INTRODUCING EVALUATION

Goals

- 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

Goals *(continued)*

- Discuss some of the practical challenges of doing evaluation
- Through case studies, illustrate how methods discussed in Chapters 8, 9, and 10 are used in evaluation, and describe some methods that are specific to evaluation
- Provide an overview of methods that are discussed in detail in the next two chapters

Why, what, where, and when to evaluate

Iterative design and evaluation is a continuous process that examines:

Why: To check users' requirements and confirm that users can utilize the product and that they like it

What: A conceptual model, early and subsequent prototypes of a new system, more complete prototypes, and a prototype to compare with competitors' products

Where: In natural, in-the-wild, and laboratory settings

When: Throughout design; finished products can be evaluated to collect information to inform new products

Bruce Tognazzini tells you why you need to evaluate

"Iterative design, with its repeating cycle of design and testing, is the only validated methodology in existence that will consistently produce successful results. If you don't have user-testing as an integral part of your design process you are going to throw buckets of money down the drain."

See AskTog.com for topical discussions about design and evaluation

Types of evaluation

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

Living labs

- 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
- 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 *(continued)*

- More recent examples include whole blocks and cities that house hundreds of people, for example, Verma et al., research in Switzerland (2017)
- Many citizen science projects can also be thought of as living labs, for instance, 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

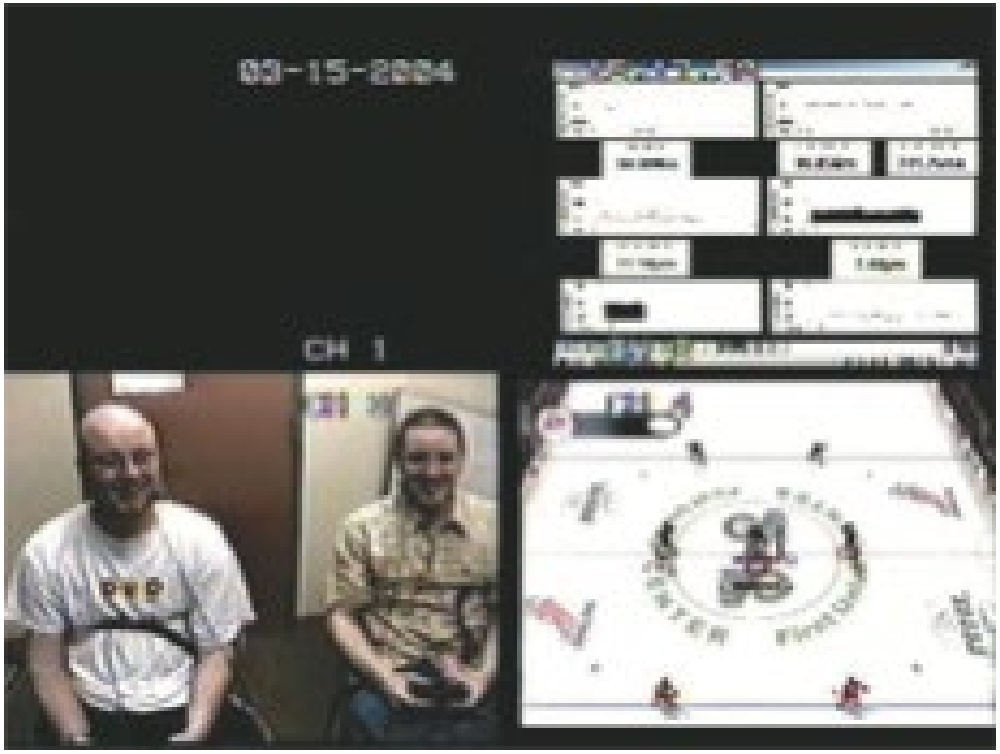
Evaluation case studies

- A classic experimental investigation into the physiological responses of players of a computer game
- 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

Challenge and engagement in a collaborative immersive game

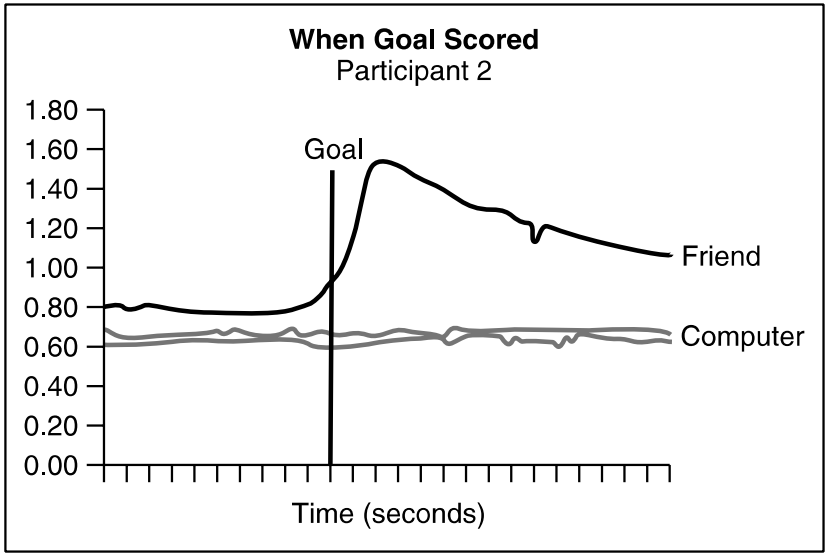
- Physiological measures were used
- Players were more engaged when playing against another person than when playing against a computer
- Why was the physiological data collected normalized?

Physiological data of participants in a videogame

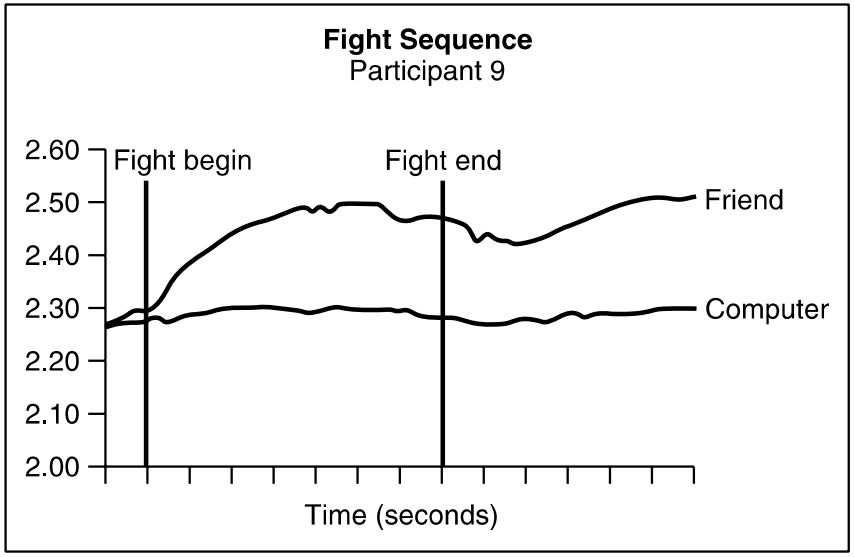


Source: Mandryk and Inkpen (2004), "*The Physiological Indicators for the Evaluation of Co-located Collaborative Play,"* CSCW'2004, pp 102-111. Reproduced with permission of ACM Publications.

Example of physiological data



(a)



(b)

A participants' skin response when scoring a goal against a friend (a), and another participants' response when when engaging in a hockey fight against a friend versus against the computer (b).

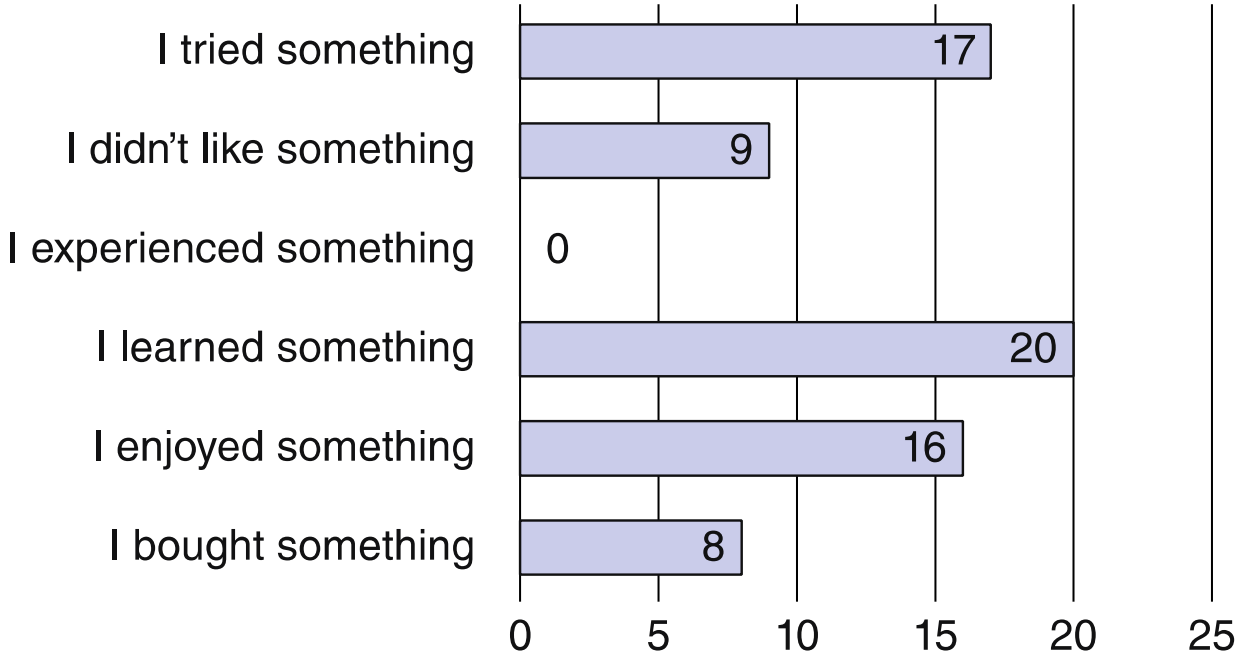
Source: Mandryk and Inkpen (2004), "*The Physiological Indicators for the Evaluation of Co-located Collaborative Play,"* CSCW'2004, pp 102-111. Reproduced with permission of ACM Publications.

Ethnobot app used at the Royal Highland Show

- The Ethnobat 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 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.

What did we learn from the case studies?

- How to observe users in the lab and in natural settings
- How evaluators excerpt different levels of control in the lab and in natural settings and in crowdsourcing evaluation studies
- Use of different evaluation methods

What did we learn from the case studies? *(continued)*

- How to develop different data collection and analysis techniques to evaluate user experience goals such as challenge and engagement
- The ability to run experiments on the Internet that are quick and inexpensive using crowdsourcing
- How a large number of participants can be recruited using 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
- Ecological validity
- Expert review or criticism
- Field study
- Formative evaluation
- Heuristic evaluation
- Informed consent form
- In the wild evaluation
- Living laboratory
- Predictive evaluation
- Reliability
- Scope
- Summative evaluation
- Usability laboratory
- User studies
- Usability testing
- Users or participants
- Validity

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 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 high authority, such as the 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?

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 laboratories, less controlled field settings, or where users are not present

Summary *(continued)*

- 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
- 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