

Chapter 13

WHAT IS INTERACTION DESIGN?

The aims

- Explain the key concepts and terms used in evaluation
- Introduce 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 mix and modify methods to meet the demands of evaluating novel systems.
- Discuss some of the challenges that evaluators have to consider when doing evaluation.
- Illustrate how methods discussed in Chapters 7 and 8 are used in evaluation and describe some methods that are specific to evaluation.

Why, what, where and when to evaluate

Iterative design & evaluation is a continuous process that examines:

- Why: to check users' requirements and that they 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.

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 involving users, eg usability testing & experiments in laboratories and living labs.
- Natural settings involving users, eg field studies and in the wild studies to see how the product is used in the real world.
- Settings not involving users, e.g. to predict, analyze & model aspects of the interface analytics.

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.
- Eg the Aware Home was embedded with a complex network of sensors and audio/video recording devices (Abowd et al., 2000).

Usability testing & field studies can compliment

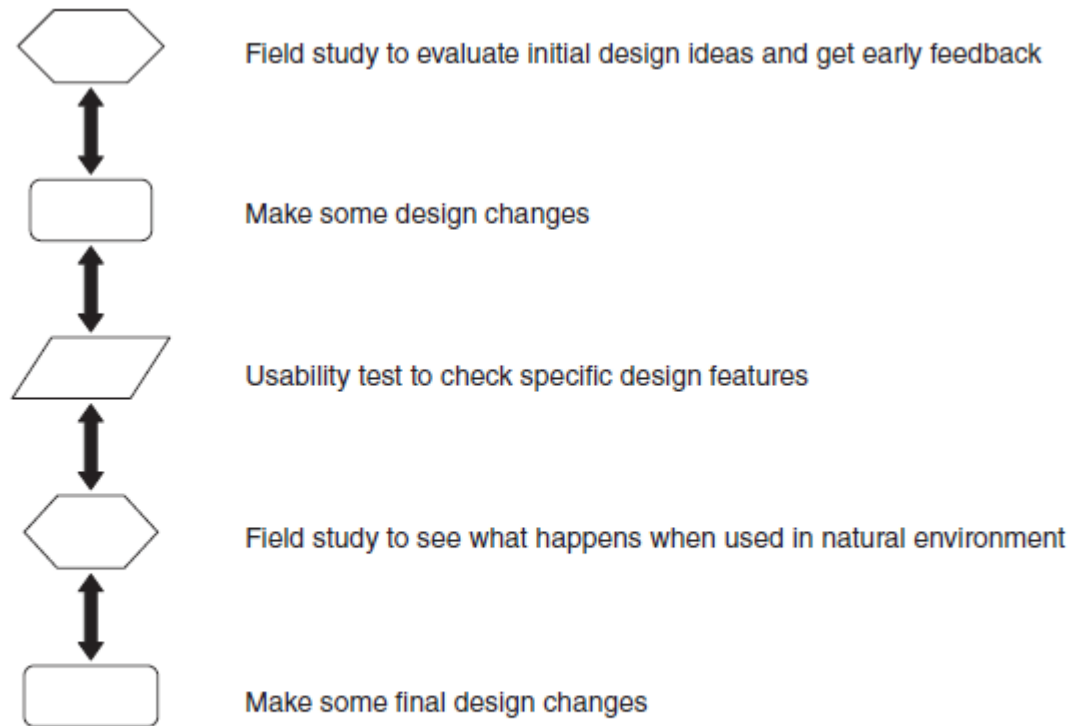


Figure 13.1 Example of the way laboratory-based usability testing and field studies can complement each other

Evaluation case studies

- Experiment to investigate a computer game
- In the wild field study of skiers
- Crowdsourcing

Challenge & 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.
- What precautionary measures did the evaluators take?

Challenge & engagement in a collaborative immersive game

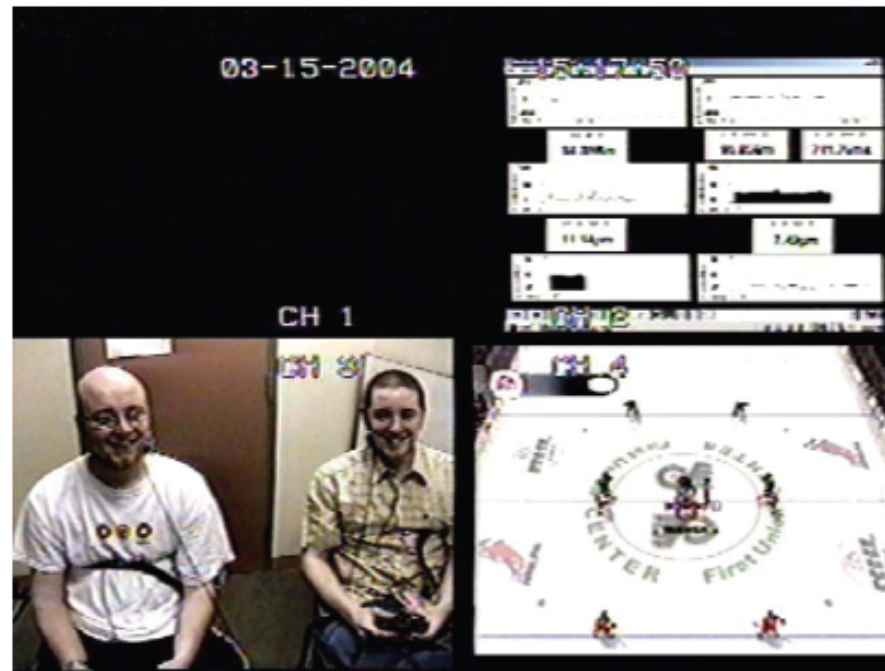


Figure 13.2 The display shows the physiological data (top right), two participants, and a screen of the game they played

Source: Mandryk and Inkpen (2004) Physiological Indicators for the Evaluation of Co-located Collaborative Play, CSCW'2004, pp. 102–111. ©2004 Association for Computing Machinery, Inc. Reprinted by permission.

What does this data tell you?

	Playing against computer		Playing against friend	
	Mean	St. Dev.	Mean	St. Dev.
Boring	2.3	0.949	1.7	0.949
Challenging	3.6	1.08	3.9	0.994
Easy	2.7	0.823	2.5	0.850
Engaging	3.8	0.422	4.3	0.675
Exciting	3.5	0.527	4.1	0.568
Frustrating	2.8	1.14	2.5	0.850
Fun	3.9	0.738	4.6	0.699

Table 13.1 Mean subjective ratings given on a user satisfaction questionnaire using a five-point scale, in which 1 is lowest and 5 is highest for the 10 players. Identifying strongly with an experience state is indicated by a higher mean. The standard deviation indicates the spread of the results around the mean. Low values indicate little variation in participants' responses, high values indicate more variation

Why study skiers in the wild ?



(a)



(b)

Figure 13.4 (a) A skier wearing a helmet with an accelerometer (dark red box) and a mini-camera (black cylinder) placed on it for assessing the skier's performance and (b) the smartphone that provides feedback to the skier in the form of visualizations

Source: Jambon and Meillon (2009) User experience in the wild. In: *Proceedings of CHI '09*, ACM Press, New York, p. 4070.

e-skiing system components

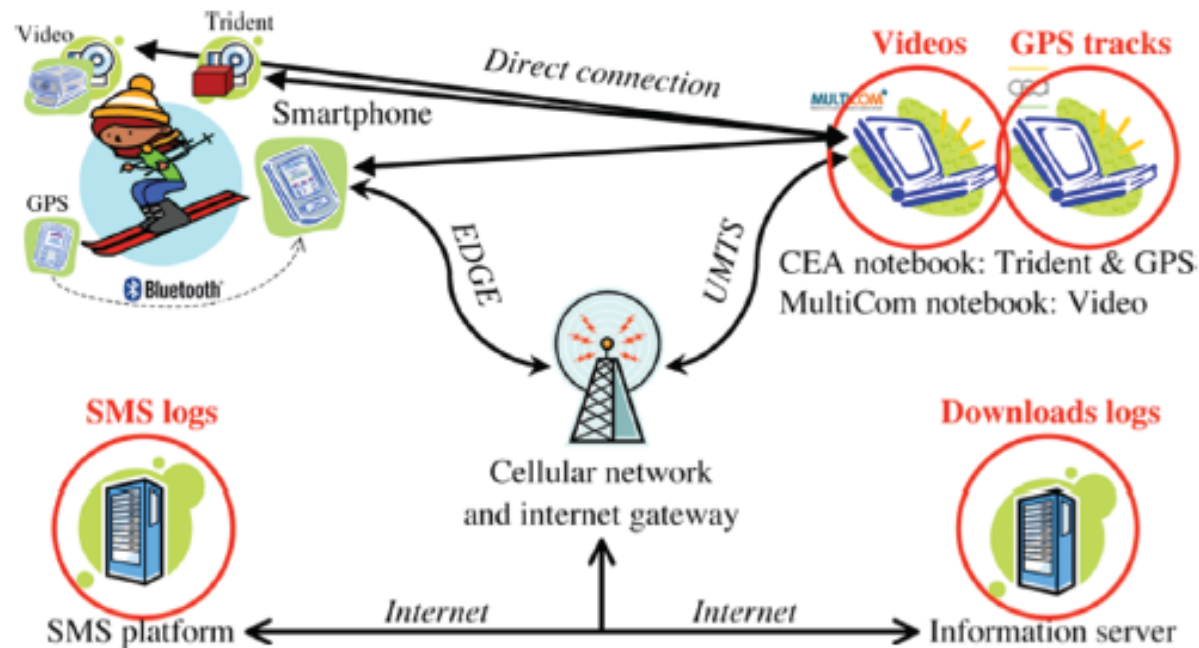


Figure 13.5 Components of the e-skiing system. Back arrows indicate the data transfers between devices, servers, and linking systems. Arrow shapes indicate different types of communications and the red circles indicate the data collection points

Source: Jambon *et al* (2009) User experience in the wild. In: *Proceedings of CHI '09*, ACM Press, New York, p. 4070.

What did we learn from the case studies?

- How to observe users in natural settings.
- Unexpected findings resulting from in the wild studies.
- Having 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 to recruit a large number of participants using Mechanical Turk. Test text

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 crit

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 their rights.
- Informed consent forms provide this information.
- 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 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, e.g. observation, interviews, and questionnaires.
- Evaluations can be done in controlled settings such as laboratories, less controlled field settings, or where users are not present.
- 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.