

Introducing Evaluation

Lecture 13 Dr Kristina Lapin



Outline

- The types of evaluation
- Evaluation case studies
- Evaluation framework DECIDE
- Language of evaluation

The aims

- Explain the key concepts used in evaluation.
- Introduce different evaluation methods.
- Show how different methods are used for different purposes at different stages of the design process and in different contexts.
- Show how evaluators mix and modify methods.
- Discuss the practical challenges
- 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 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.

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.

<http://www.asktog.com/columns/037TestOrElse.html>

The language of evaluation

- Analytical evaluation
- Controlled experiment
- Field study
- Formative evaluation
- Heuristic evaluation
- Predictive evaluation
- Summative evaluation
- Usability laboratory
- User studies
- Usability studies
- Usability testing
- User testing

The types of evaluation

- Controlled settings involving users
 - Examples
 - laboratories and living labs
 - Methods:
 - usability testing
 - experiments
 - User's activities are controlled in order to test hypothesis and measure or observe certain behaviors
- + Good at revealing usability problems
- Poor at capturing context of use

Usability lab



- A combination of methods
 - experiments,
 - observation,
 - interviews
 - questionnaires
- Controlled environment

http://iat.ubalt.edu/?page_id=13

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).
 - MIT Living Labs have been developed to evaluate people's everyday lives
livinglabs.mit.edu.

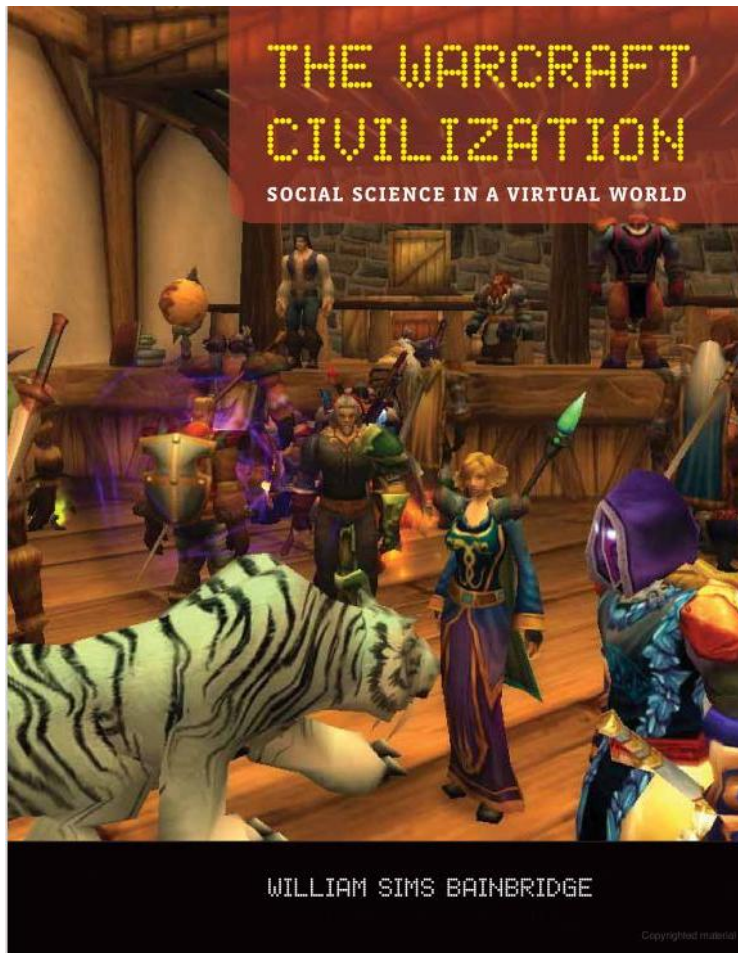
The types of evaluation

- Natural settings involving users
 - E.g. online communities and public places
 - A little or no control of users' activities in order to determine how the product would be used in the real world.
 - Method:
 - field studies to see how the product is used in the real world
- + Good at demonstrating how people use technologies
- Expensive and difficult to conduct

Natural Settings Involving Users

- Help identify opportunities for a new technologies
- Help establish requirements for a new design
- Facilitate the introduction of technology, or inform deployment of existing technology in the new context
- Methods:
 - observation and logging
 - In the wild studies: real and virtual environments

Studies in the wild



- ethnographic participant observation for two years 2007-2009
- Academic conference in WoW



The types of evaluation

- Any settings not involving users
 - consultants critique
 - to predict, analyze & model aspects of the interface analytics
 - Methods:
 - Inspections, heuristics, walkthroughs, models and analytics
- + Cheap and quick to perform
- Can miss unpredictable usability problems and subtle aspects of user experience

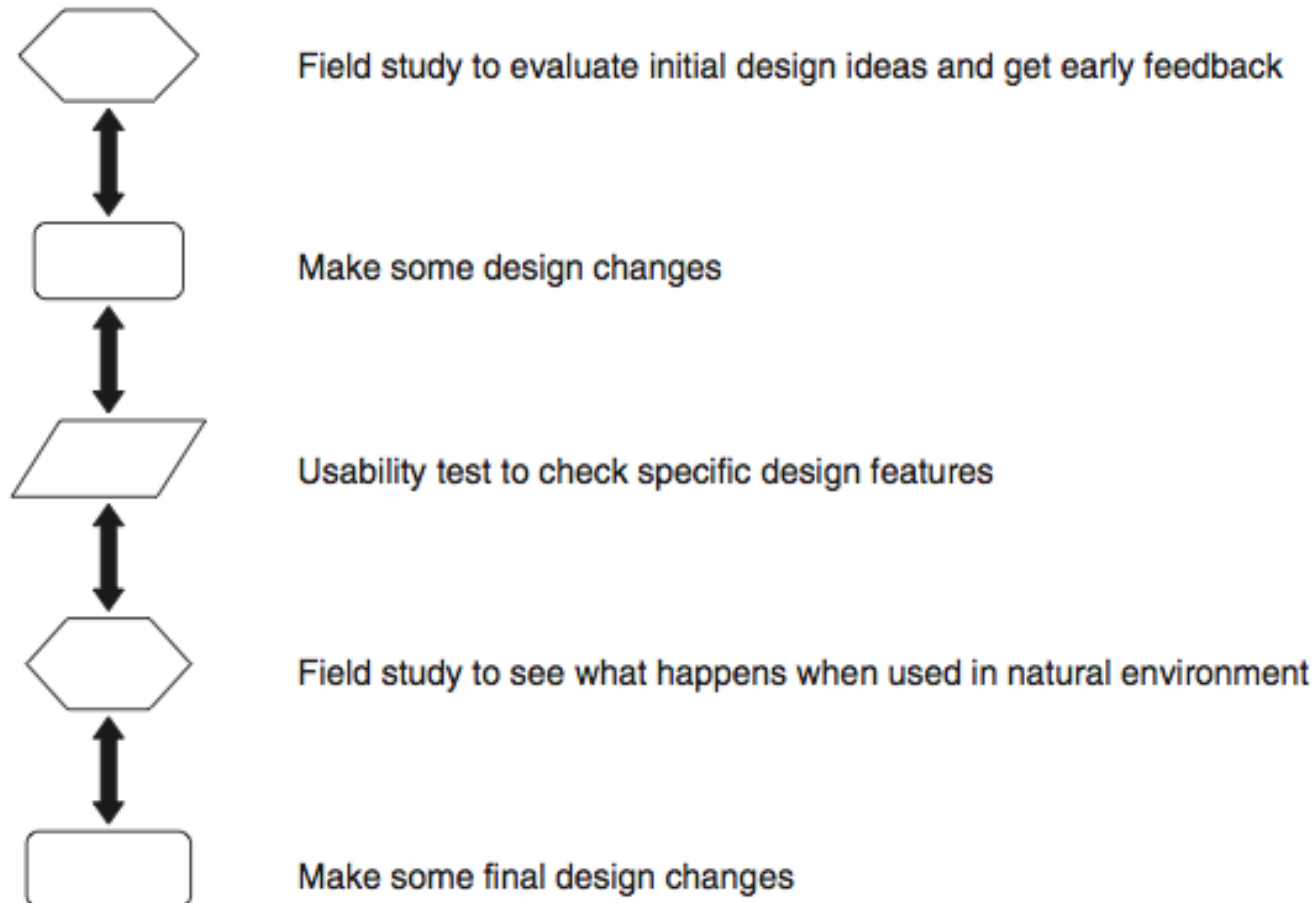
Any settings Not Involving Users

- Inspection methods or modelling predict user user bahavior and to identify usability problems
- Heuristic evaluation (Nielsen, Tahir, 2002)
- Cognitive Walkthrough ([Warthon, Rieman, lewis, Polson 1994](#))
- Analytics – logging data analysis ([Arikan, 2008](#))
- Models – for comparing efficacy
 - Keyboard Level Models, Fitts' Law, Hick's Law

Characteristics of approaches

	Controlled environme nt with users	Natural environme nt with users	Any setting without users
Users	do task	natural	not involved
Location	controlled	natural	anywhere
When	prototype	early	prototype
Data	quantitative	qualitative	problems
Feed back	measures & errors	descriptions	problems
Type	applied	naturalistic	expert

Usability testing & field studies can compliment



Opportunistic evaluations

- Done early in the design process to provide designers with feedback quickly about the design idea.
- Early evaluations are informal and cheap.
- Helps developers to decide if an idea needs to be modified or abandoned

Outline

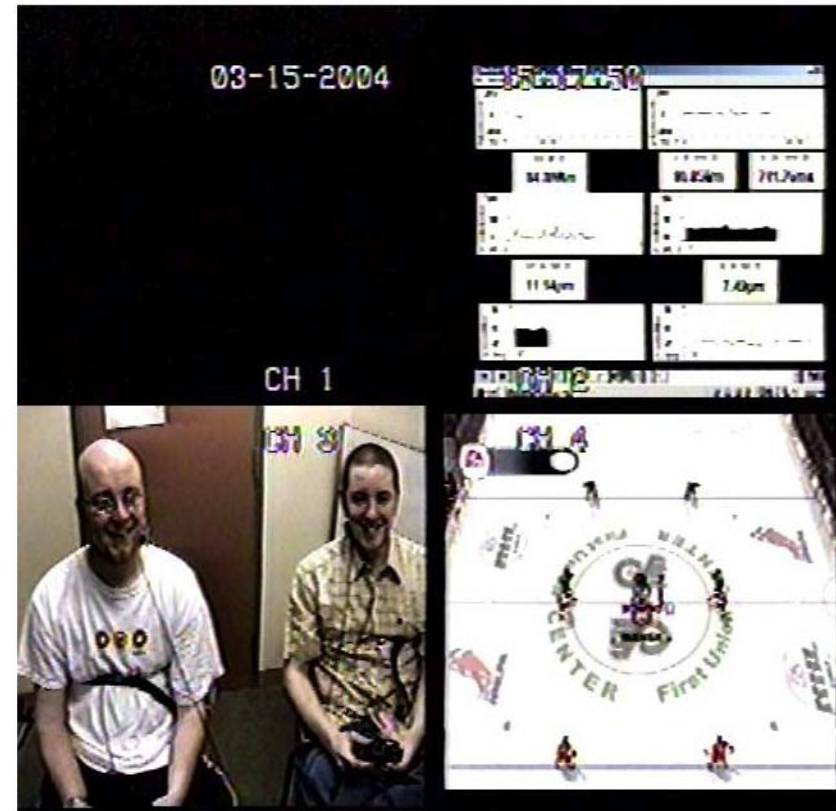
- The types of evaluation
- Evaluation case studies
- Evaluation framework DECIDE
- Language of evaluation

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?

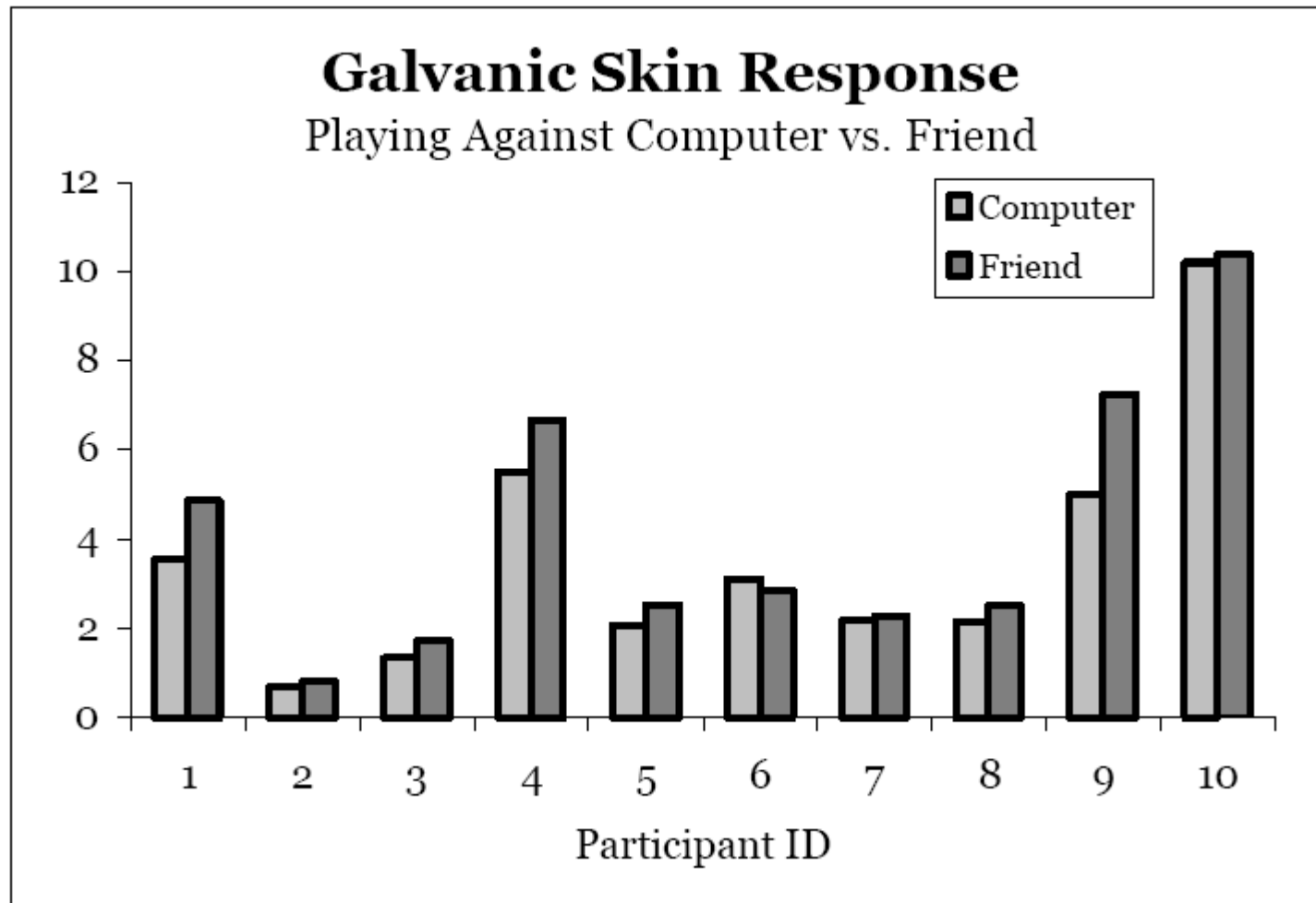


Mandryk, Inkpen 2004)

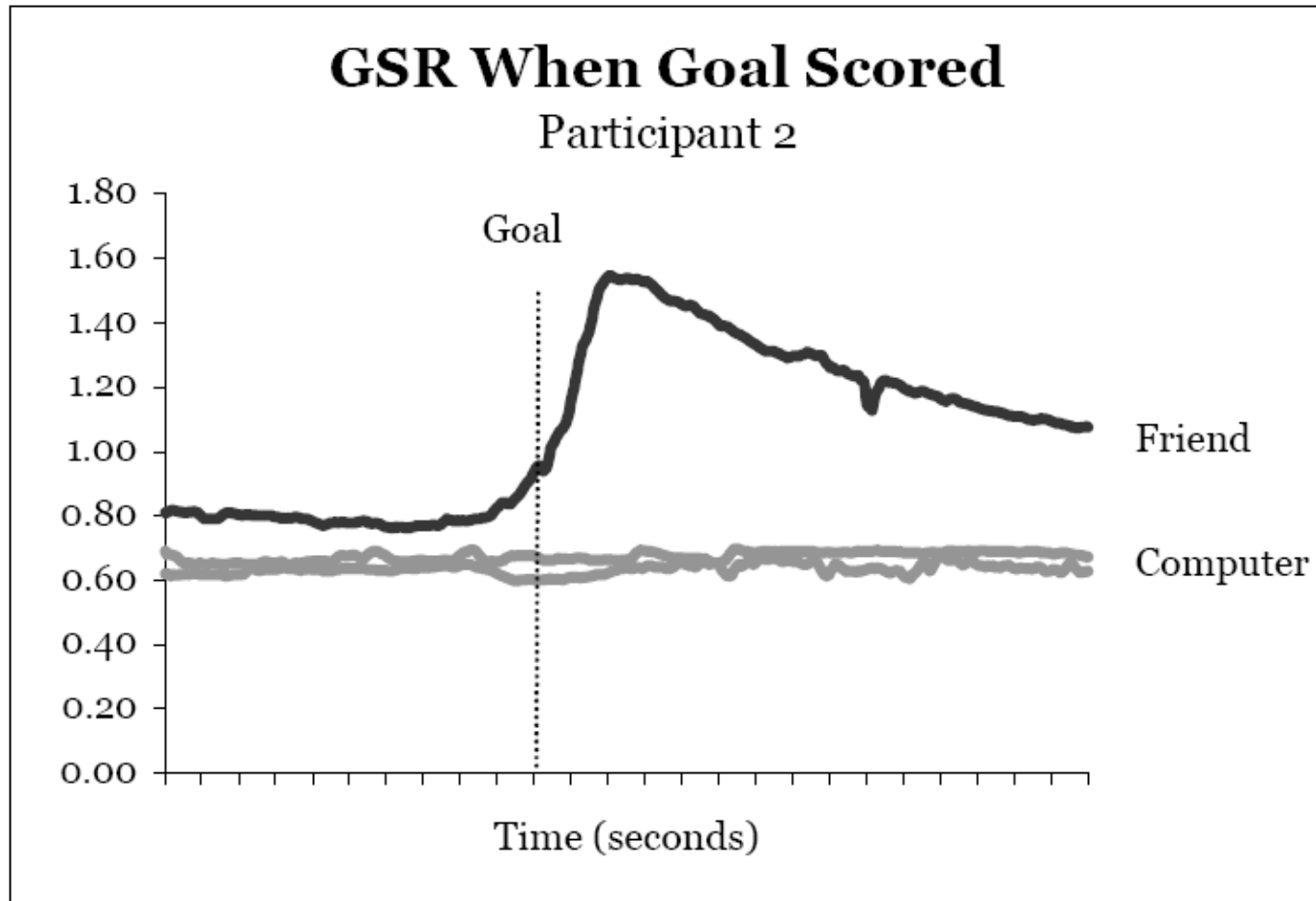
Challenge & engagement in a collaborative immersive game

	Playing against computer		Playing against friend		Difference between conditions	
	<i>Mean</i>	<i>St.Dev</i>	<i>Mean</i>	<i>St.Dev</i>	χ^2	<i>p</i>
Boring	2.3	.949	1.7	.949	4.0	.046
Challenging	3.6	1.08	3.9	.994	1.8	.180
Easy	2.7	.823	2.5	.850	1.0	.317
Engaging	3.8	.422	4.3	.675	4.0	.046
Exciting	3.5	.527	4.1	.568	6.0	.014
Frustrating	2.8	1.14	2.5	.850	.67	.414
Fun	3.9	.738	4.6	.699	6.0	.014

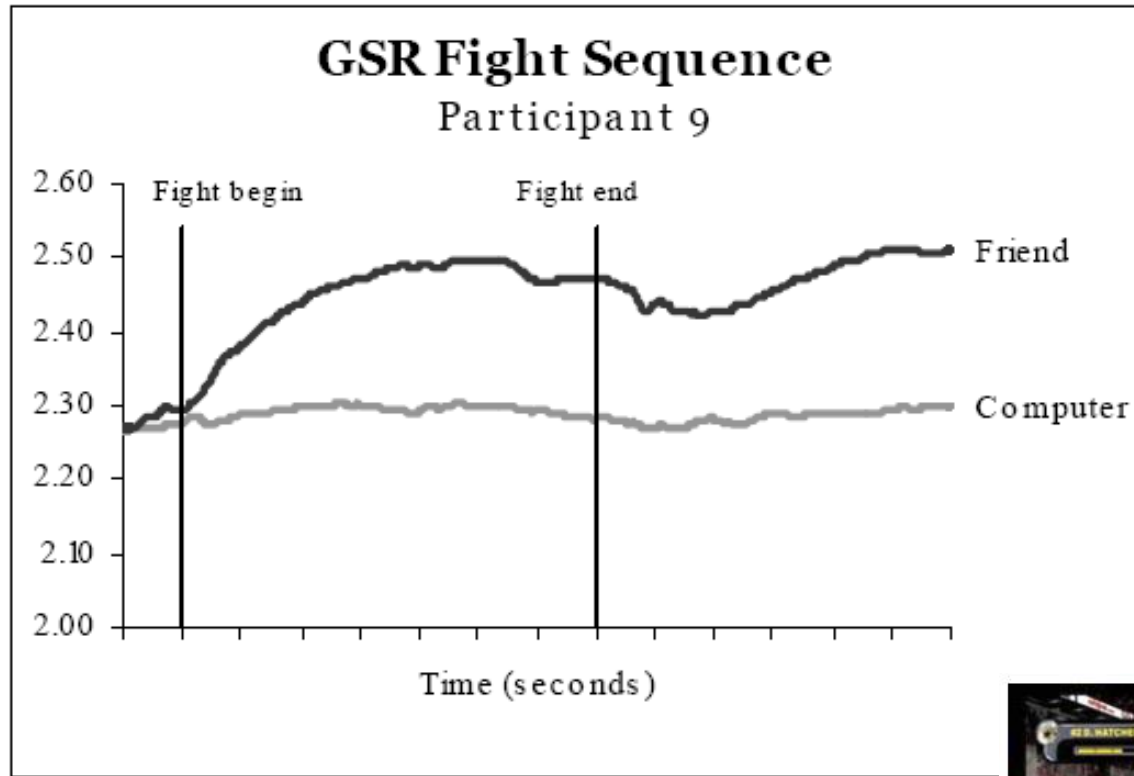
Challenge & engagement in a collaborative immersive game



Challenge & engagement in a collaborative immersive game



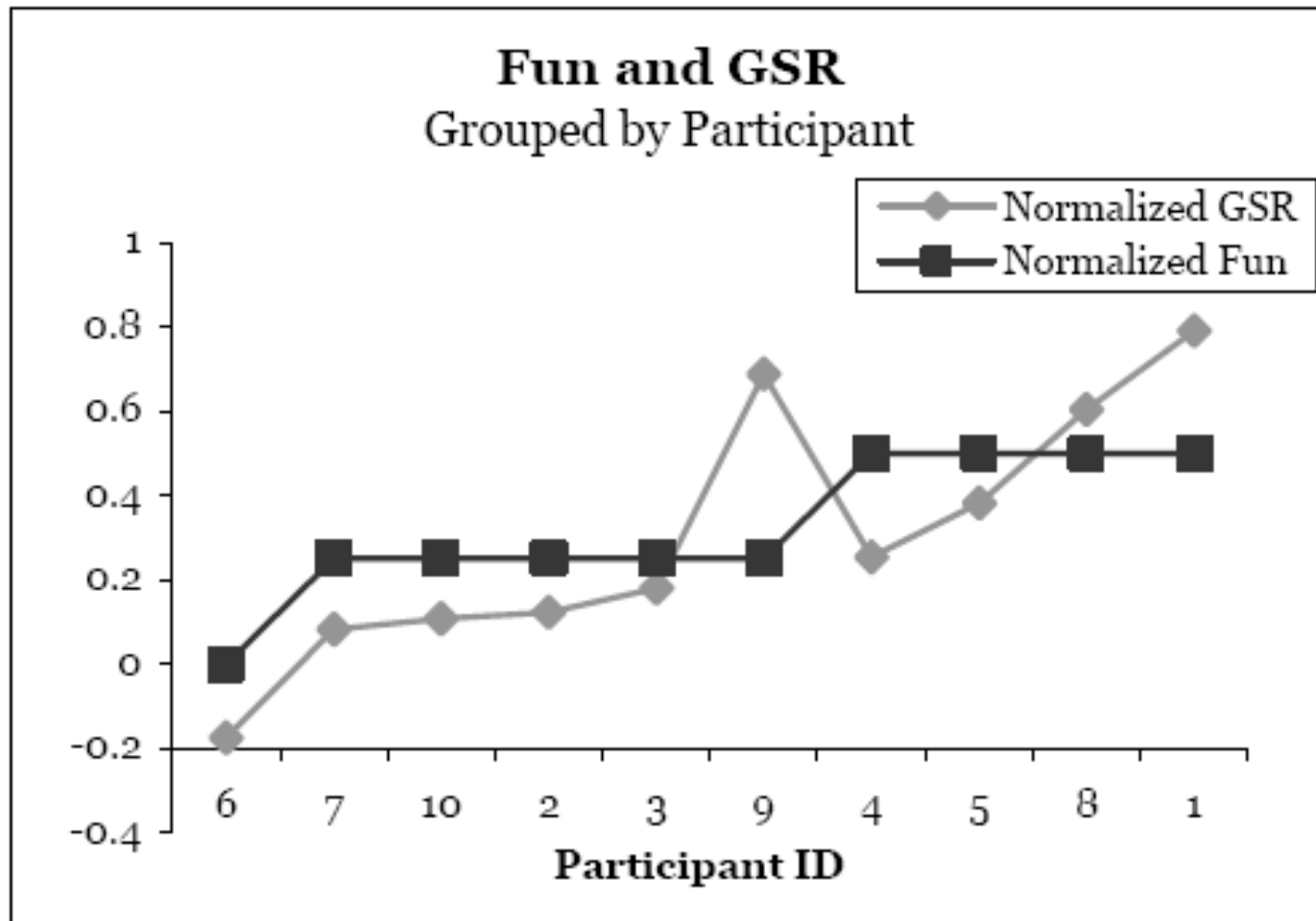
Challenge & engagement in a collaborative immersive game



Mandryk, Inkpen 2004)



Challenge & engagement in a collaborative immersive game



Challenge & engagement in a collaborative immersive game

- What kind of setting was used in this experiment?
- How much control did the evaluators exert?
- Which methods were recorded and when?

Why study skiers in the wild?



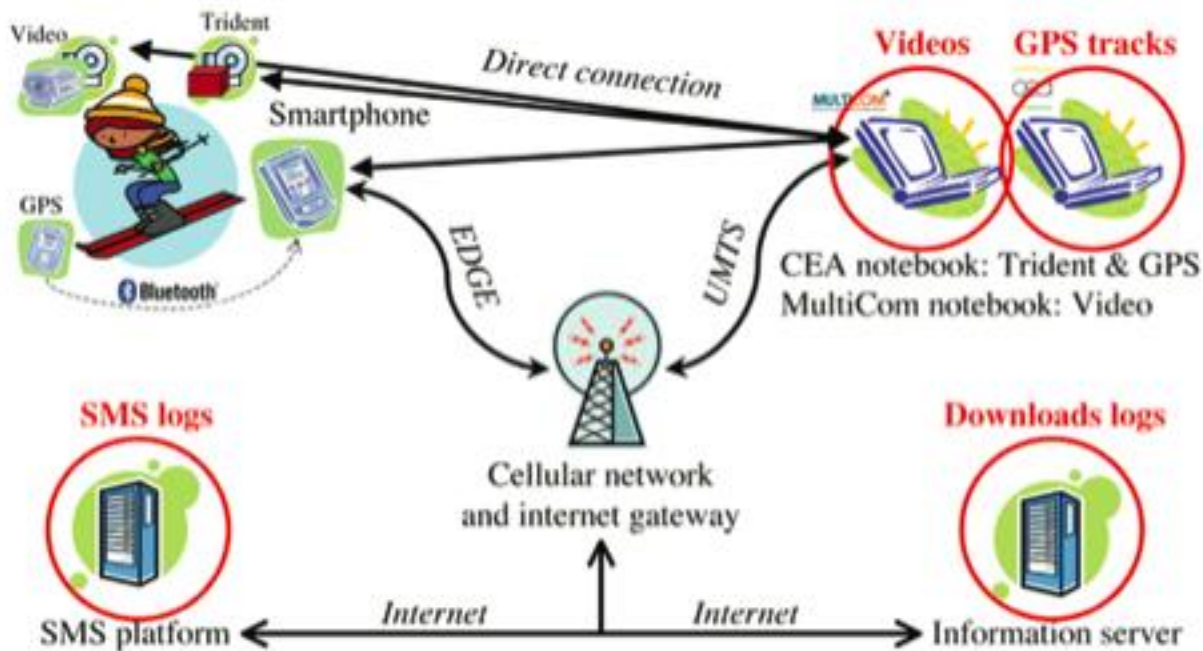
Jambon et al. (2009) User experience in the wild. In: Proceedings of CHI '09, ACM Press, New York, p. 4070-4071.

2014.04.08 11:50

Adapted from www.id-book.com

<#>

e-skiing system components



Jambon et al. (2009) *User experience in the wild*. In: *Proceedings of CHI '09*, ACM Press, New York, p. 4072.

Crowdsourcing-when might you use it?

amazon mechanical turk
Artificial Intelligence

Your Account

HITs

Qualifications

Already have an account?
Sign in as a [Worker](#) | [Requester](#)

[Introduction](#) | [Dashboard](#) | [Status](#) | [Account Settings](#)

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

161,325 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

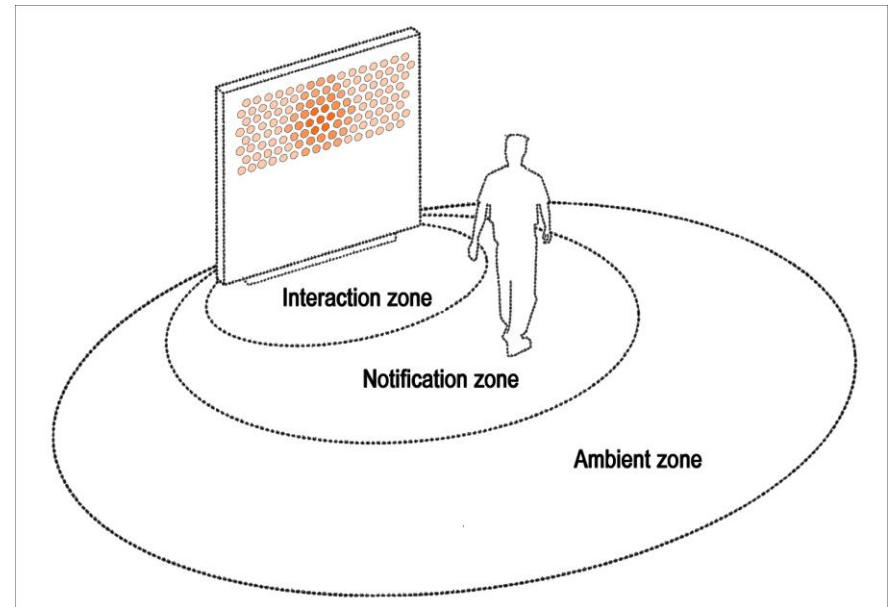
As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



Evaluating an ambient system

- The Hello Wall is a new kind of system that is designed to explore how people react to its presence.
- What are the challenges of evaluating systems like this?



Evaluation methods

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

Outline

- The types of evaluation
- Evaluation case studies
- Evaluation framework DECIDE
- Language of evaluation

The aims are:

- Introduce and explain the DECIDE framework.
- Discuss the conceptual, practical, and ethical issues involved in evaluation.

DECIDE: a framework to guide evaluation

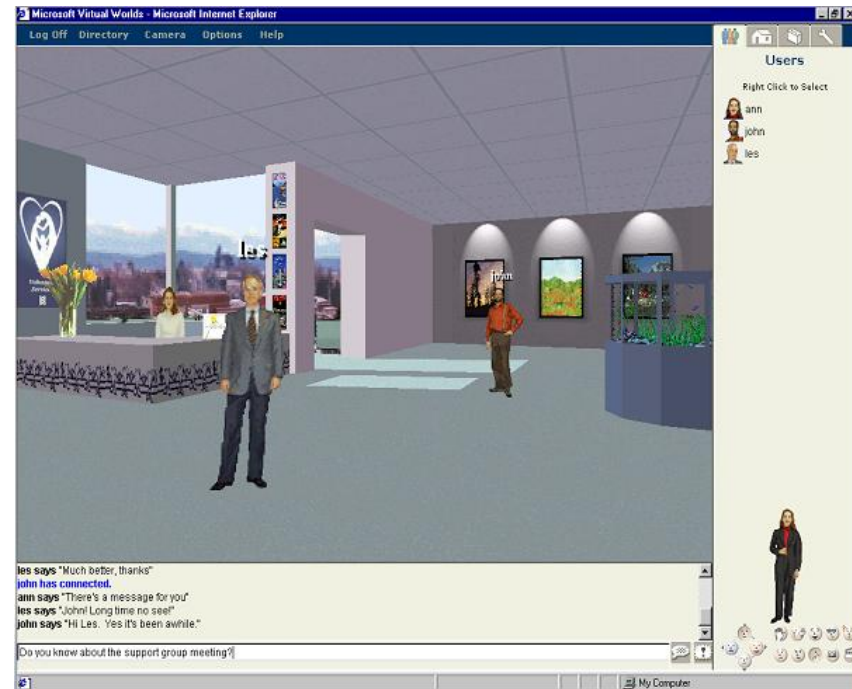
- **D**etermine the *goals*.
- **E**xplore the *questions*.
- **C**hoose the evaluation *methods*.
- **I**dentify the *practical issues*.
- **D**ecide how to deal with the *ethical issues*.
- **E**valuate, analyze, interpret and present the *data*.

Determine the goals

- What are the high-level goals of the evaluation?
- Who wants it and why?
- The goals influence the methods used for the study.
- Goals vary and could be to:
 - identify the best metaphor for the design
 - check that user requirements are met
 - check for consistency
 - investigate how technology affects working practices
 - improve the usability of an existing product

1. Determine the goals

- The HutchWorld patient support system
 - distributed virtual community for the Fred Hutchinson Cancer Research Center in Seattle, Wa
- Which metaphor?

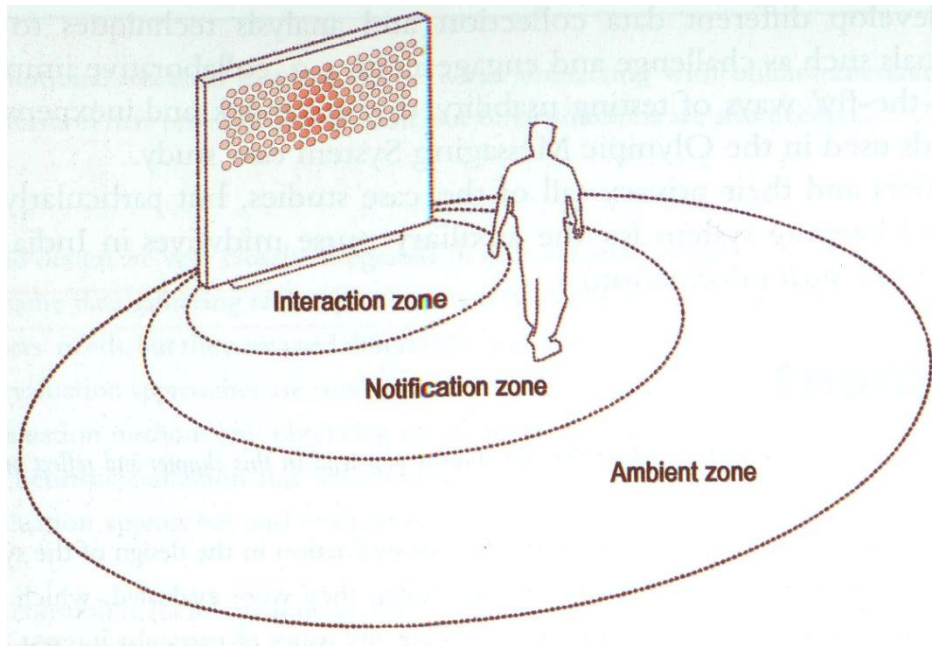


Explore the questions

- Questions help to guide the evaluation.
- The goal of finding out why some customers prefer to purchase paper airline tickets rather than e-tickets can be broken down into sub-questions:
 - What are customers' attitudes to e-tickets?
 - Are they concerned about security?
 - Is the interface for obtaining them poor?
- What questions might you ask about the design of a cell phone?

What goals and explore questions would you set for Hello.Wall?

Ambient display Hello.Wall



http://www.youtube.com/watch?v=QHNA_9i8I9I&feature=Playlist&p=C9F2937C5CF2DD51&index=2

Choose the evaluation approach & methods

- The evaluation method influences how data is collected, analyzed and presented.
- E.g. field studies typically:
 - Involve observation and interviews.
 - Involve users in natural settings.
 - Do not involve controlled tests.
 - Produce qualitative data.

Identify practical issues

For example, how to:

- Select users
- Find evaluators
- Select equipment
- Stay on budget
- Stay on schedule

Decide about ethical issues

- Develop an informed consent form
- Participants have a right to:
 - Know the goals of the study;
 - Know what will happen to the findings;
 - Privacy of personal information;
 - Leave when they wish;
 - Be treated politely.

Evaluate, interpret & present data

- Methods used influence how data is evaluated, interpreted and presented.
- The following need to be considered:
 - Reliability: can the study be replicated?
 - Validity: is it measuring what you expected?
 - Biases: is the process creating biases?
 - Scope: can the findings be generalized?
 - Ecological validity: is the environment influencing the findings? i.e. Hawthorn effect.

Key points

- Many issues to consider before conducting an evaluation study.
- These include: goals of the study; involvement or not of users; the methods to use; practical & ethical issues; how data will be collected, analyzed & presented.
- The DECIDE framework provides a useful checklist for planning an evaluation study.

Outline

- The types of evaluation
- Evaluation case studies
- Evaluation framework DECIDE
- Language of evaluation

The language of evaluation

Analytics

Analytical
evaluation

Controlled
experiment

Expert review or crit

Field study

Formative
evaluation

Heuristic evaluation

In the wild
evaluation

Living laboratory

Predictive evaluation

Summative
evaluation

Usability laboratory

User studies

Usability testing

Users or participants

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.

References

- Rogers, Sharp, Preece (2011). [Interaction design](#): Beyond Human Computer Interaction. Wiley.
- G.D. Abowd, C.G. Atkeson, A.E. Bobick, I.A. Essa, B. MacIntyre, E.D. Mynatt, T.E. Starner (2000) Living Laboratories: The Future Computing Environments Group at the Georgia Institute of Technology. In [CHI'00 Extended Absrtracts on Human Factors in Couputing Systems, CHI 2000, ACM, pp. 215-206](#)
- J. Nielsen, M. Tahir (2002) Homepage usability : 50 websites deconstructed. New Riders Press. (MIF bibliotekoje)
- Jeffrey Heer, Michael Bostock (2010) Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In [Proceedings of CHI 2010, ACM, pp. 203-212.](#)
- Bainbridge, W.S. (2010) *The Warcraft Civilization: Social Science in Virtual World*. [MIT Press, Cambridge, MA.](#)

References

- Cockton, G., Woolrych, A. (2008) Inspection-Based Evaluations. In *The Human-Computer Interaction handbook: Fundamentals, \evolving technologies and Wemerging Applications*, A. Sears, J.A.Jacko (Eds) CRC Press. (MIF bibliotekoje)
- [Warthon, C., Rieman, J., Lewis, C., Polson, P \(1994\)](#) The cognitive walkthrough method: a practitioner's guide. In *Book, Usability inspection methods*, John Wiley & Sons, Inc. New York, NY, USA.
- Arikon, A. (2008). [Multichannel Marketing: metrics and methods for on- and offline seccess](#). Sybex.
- R.L. Mandryk, K.M. Inkpen (2004) Physiological Indicators for the Evaluation of Co-located Collaborative Play. In *CSCW 2004*, [ACM Press, pp. 102-111](#).
- T. Hollingsed, D.G. Novick (2007) Usability Inspection Methods after 15 Years of Research and Practice. [SIGDOC'07, ACM](#).

A project for you ...

- “The Butterfly Ballot: Anatomy of disaster” was written by Bruce Tognazzini, and you can find it by going to AskTog.com and looking through the 2001 column.
- Alternatively go directly to:
<http://www.asktog.com/columns/042ButterflyBallot.html>

A project for you ... continued

- Read Tog's account and look at the picture of the ballot card.
- Make a similar ballot card for a class election and ask 10 of your friends to vote using the card. After each person has voted ask who they intended to vote for and whether the card was confusing. Note down their comments.
- Redesign the card and perform the same test with 10 different people.
- Report your findings.

A project for you ...

- Find an evaluation study from the list of URLs on this site or one of your own choice.
- Use the DECIDE framework to analyze it.
- Describe the aspects of DECIDE that are explicitly addressed in the report and which are not.
- On a scale of 1-5, where 1 = poor and 5 = excellent, how would you rate this study?