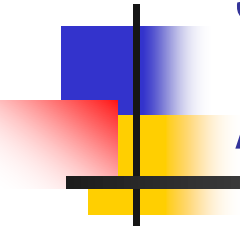


SOEN 6751: Human Computer Interface Design

Introducing Evaluation

Based on Chapter 14 of the textbook
and the slides from Prof. T. Radhakrishnan

What is UI evaluation?

- Test your design ideas with representative end-users in as natural an environment as possible
  - The more realistic the test environment and the more representative the users, the more accurate the results
- Discover how good/bad your design is in as many ways as possible
  - Important to communicate this to the team beforehand
- The opportunity to ask end-users about the why behind their actions and opinions with respect to your design

What is UI evaluation? ...

- A de-risking activity and a diagnostic tool
  - Check that design assumptions are valid
  - Detect and fix usability issues early in the cycle
  - Identify potential issues to monitor moving forward
- Make informed design decisions
  - Test alternative design ideas
  - Avoid opinion wars
- Verify that you have achieved usability goals
- A dress rehearsal
  - Opportunity for stake holders, influential end-users, etc. to peak behind the curtain

What UI evaluation isn't

- UI evaluation is *not* a design review with team members or stakeholders
  - Design reviews with team members are essential
- It is *not* QA of the software!
  - If your S/W crashes, produces invalid or inconsistent results, or otherwise misbehaves then it is not ready to be presented to users
- It is *not* a rubber stamp of approval, applied on a product before it goes out the door
  - *Can you evaluate the UI quickly – we're shipping in two weeks…*

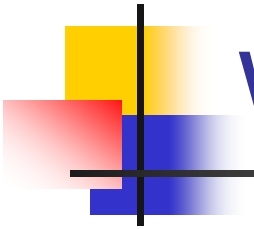
Why Evaluate the Usability of User Interface Designs?

- Does the Interface Meet the Usability Requirements?
  - Effective
  - Efficient
  - Engaging
  - Error tolerant
  - Easy to learn
- Exploring Other Concerns in Evaluations
  - Why users are unable to complete tasks easily?
  - Is the UI developed for all levels of users?
  - Are all design features acceptable to users?

Why, what, where and when to evaluate

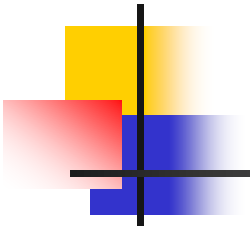
- Iterative design and evaluation is a continuous process that examines:
  - **Why**: to check that users can use the product and that they like it.
  - **What**: a conceptual model, early prototypes of a new system and later, more complete prototypes.
  - **Where**: in natural and laboratory settings.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

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

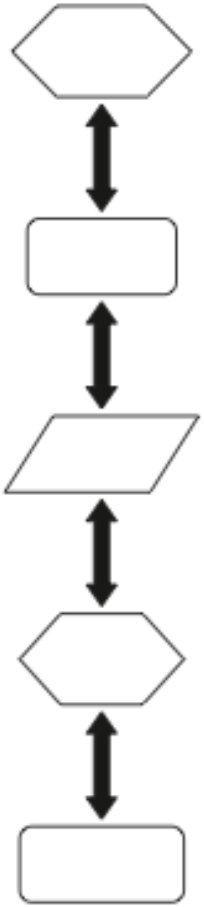
Usability lab

http://iat.ubalt.edu/usability_lab/

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.
- They are done in natural settings involving users, e.g. field studies to see how the product is used in the real world

Usability testing & field studies can compliment



Field study to evaluate initial design ideas and get early feedback

Make some design changes

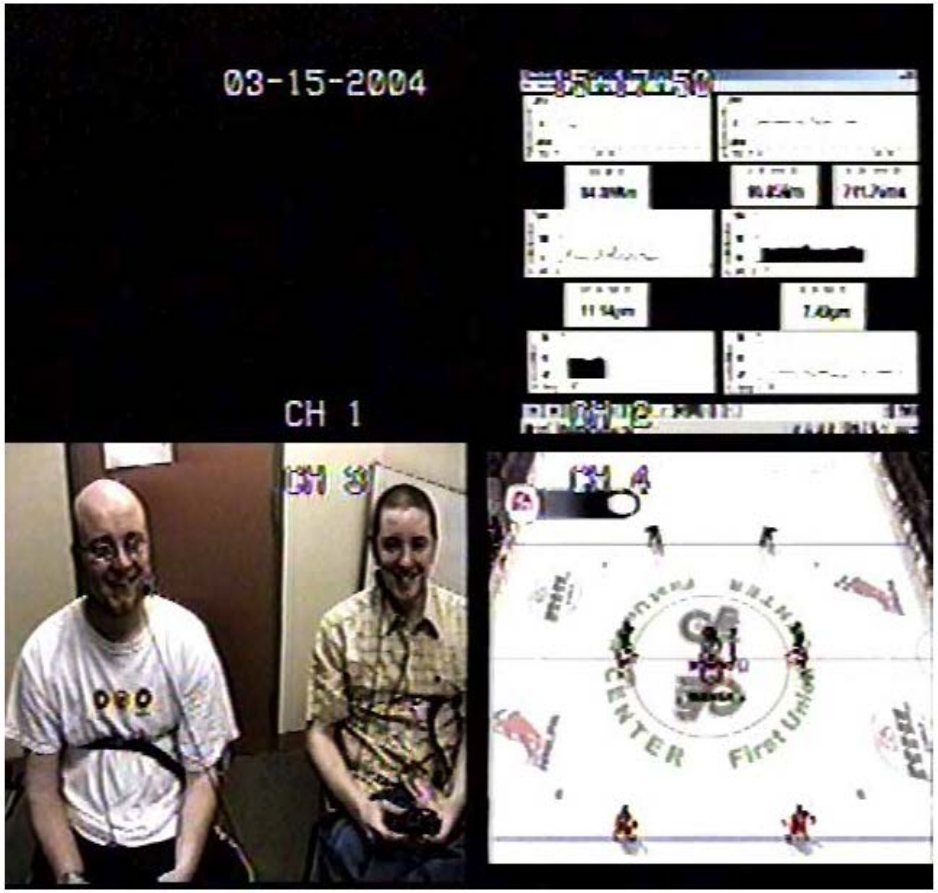
Usability test to check specific design features

Field study to see what happens when used in natural environment

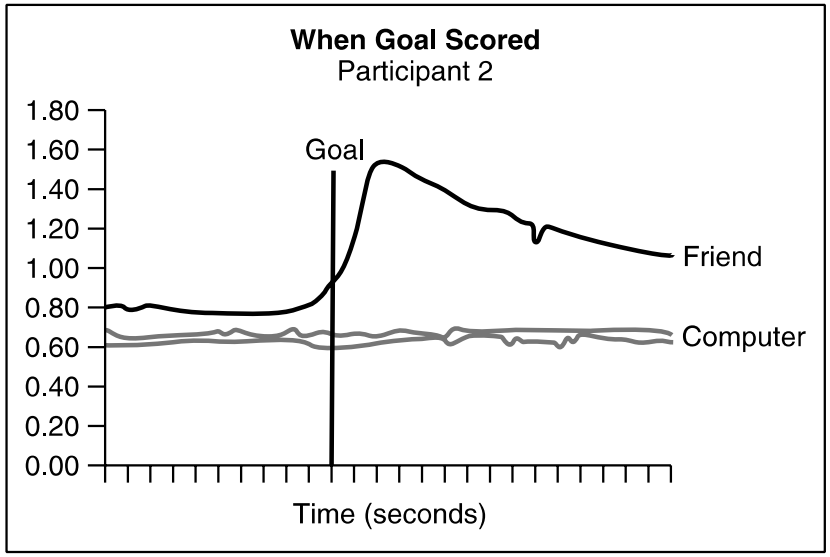
Make some final design changes

Challenge & engagement in a collaborative immersive game

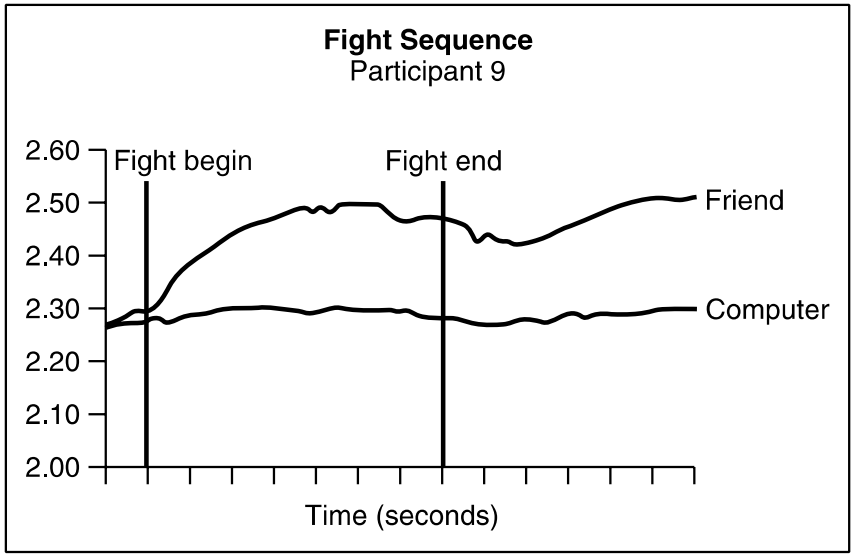
- Physiological measures were used.
- Players were more engaged when playing against another person than when playing against a computer.
- What were the precautionary measures that the evaluators had to take?



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.

What does this data tell you?

high values indicate more variation

	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

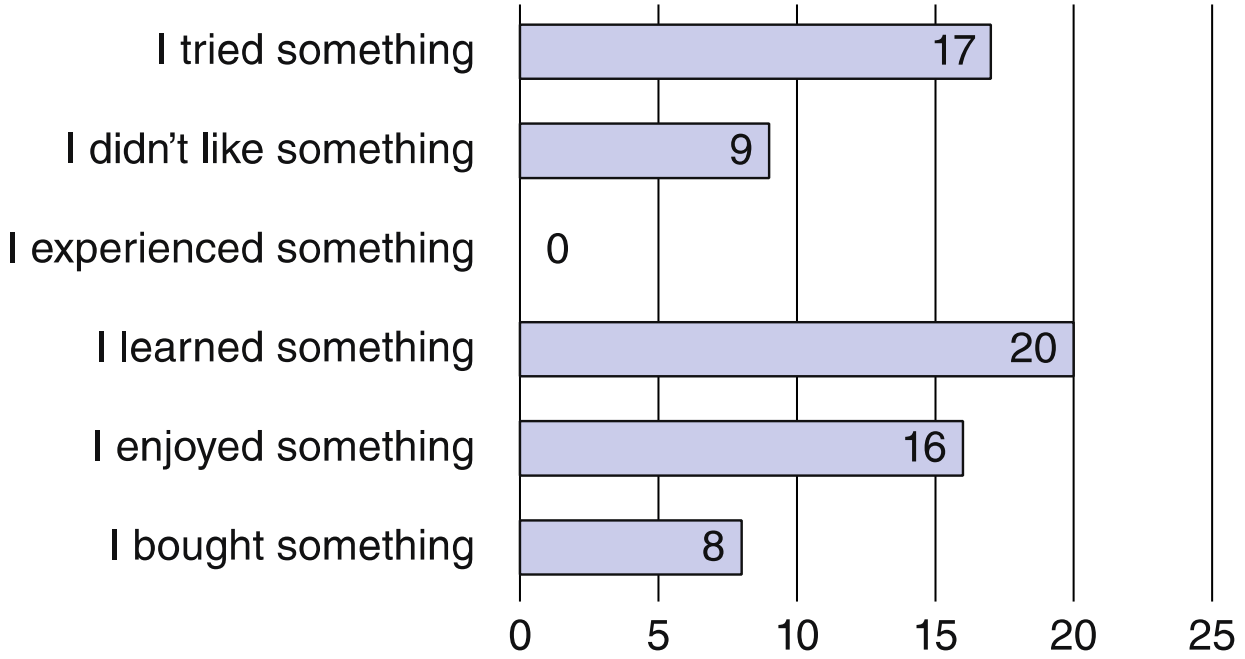
Source: Mandryk and Inkpen (2004).

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.

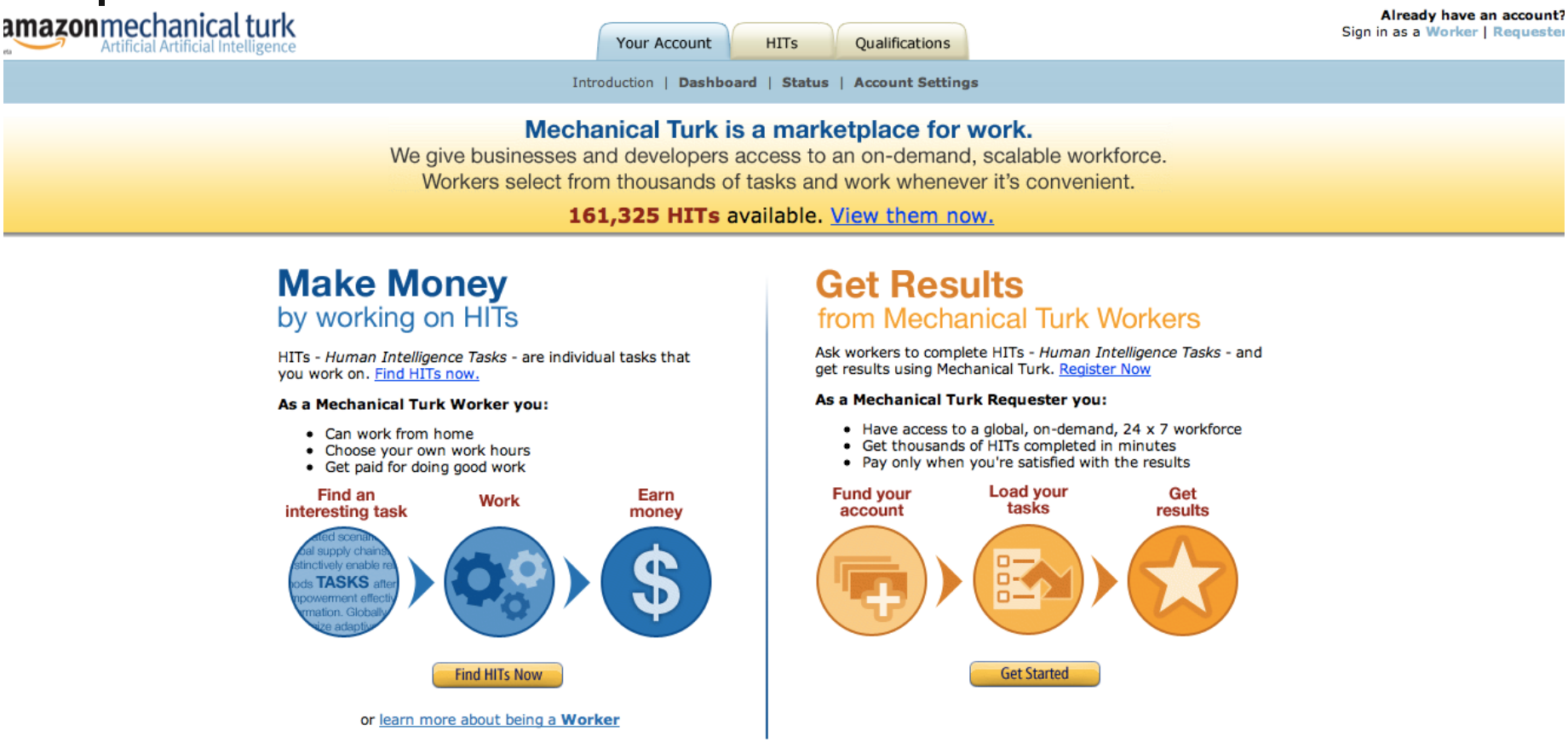
Experience responses submitted in Ethnobot



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.

Crowdsourcing- when might you use it?



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.

The language of evaluation

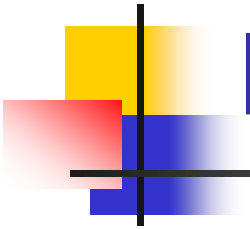
Analytics
Analytical evaluation
Biases
Controlled experiment
Crowdsourcing
Ecological validity
Expert review or crit
Field study
Formative evaluation
Heuristic evaluation
Informed consent form
In the wild evaluation
Living laboratory
Predictive evaluation
Reliability
Scope
Summative evaluation
Usability 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, e.g. 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?



Evaluation approaches

- Usability testing
- Field studies
- Analytical evaluation
- Combining approaches
- Opportunistic evaluations

Characteristics of approaches

	Usability testing	**Field studies**	**Analytical evaluation**
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

Evaluation approaches and methods

Method	Usability testing	Field studies	Analytical evaluation
Observing	x	x	
Asking users	x	x	
Asking experts		x	x
Testing	x		
Modeling			x

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 main evaluation approaches are: usability testing, field studies, and analytical evaluation.
- The main methods are: observing, asking users, asking experts, user testing, inspection, and modeling users' task performance.
- Different evaluation approaches and methods are often combined in one study.
- Triangulation involves using a combination of techniques to gain different perspectives, or analyzing data using different techniques.
- Dealing with constraints is an important skill for evaluators to develop.