

Evaluation

Interaction Design Studio (DECO I 200)

Outline

Why, what, where and when to evaluate

Some questions to ask when beginning an evaluation

Evaluation paradigms and techniques

Some different ways to go about evaluating users

DECIDE: A framework to guide evaluation

A set a guidelines for conducting evaluations

Why Evaluate?

You can't assume that everyone is like you

Just because something is obvious to you doesn't mean that it is obvious to everyone (or anyone) else

Bruce Tognazzini (Apple Employee #66):

“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.”

Tog's 5 Reasons for User Testing

1. Problems are fixed before the product is shipped, not after.
2. The team concentrates on real problems, not imaginary ones.
3. Engineers code instead of debating.
4. Time to market is sharply reduced.
5. Sales are made based on real features not promises of future upgrades.

What to Evaluate?

Different interactive products have to be evaluated differently

It is important to take time to determine what is really important to the product and so must be evaluated

Why are people going to use the product?

If someone is going to use a product what are the main things that they want to do with it?

How can performance be measured?

To evaluate a product you need to measure something (either qualitatively or quantitatively). What things can you measure about a user's interaction with a product?

What to Evaluate?

Imagine you have to think up some tests to evaluate the following products:

What are the goals of people using the products?

What can you measure to evaluate performance?

Web Browser

Traffic Control System

Cuddly Toy (for children aged 2+)

Mobile Phone Case

Community Web Site

Where to Evaluate?

Choosing a place to evaluate a product depends on what is being evaluated.

e.g. Evaluating how long it takes someone to find a piece of information on a website may be best done in a laboratory where the evaluators have the most amount of control over the conditions.

Some things can't be evaluated in a lab.

e.g. Evaluating whether a toy is robust enough to handle daily use and whether children enjoy interacting with it would be better done in a natural setting so that the evaluators can observe what children do when left to their own devices.

When to Evaluate?

Evaluation can happen throughout the design process but its purpose changes

Early in the design process evaluations are used to assess potential design ideas while later in the process they are used to check that designs still satisfy user requirements

Formative Evaluations

Formative evaluations are done during the design process to check that a product satisfies user needs

Summative Evaluations

Summative evaluations are done at the end of the design process to satisfy an external body or standard

Common Issues in Evaluation

What to do if there are not many users

How to observe users without disturbing

How to evaluate within a short timeframe

How to evaluate without a large budget

How to avoid any ethical problems

How to collect useful data

How to analyse the data to provide feedback

Evaluation Paradigms

“Quick and Dirty”

Informal evaluation for rapid feedback

Usability Testing

Measuring user performance on tasks

Field Studies

Observing users in natural settings

Predictive Evaluation

Expert predictions of typical problems

“Quick and Dirty”

“Quick and Dirty” evaluations are informal meetings between designers and users

Meetings could be organised very quickly with whoever is available at the time to give feedback

“Quick and Dirty” evaluations are common practice to get rapid feedback on ideas

Focus of evaluations is on getting some fast input from other people rather than carefully documented findings

“Quick and Dirty” evaluations can be done at any stage of the design process

Early meetings might discuss general ideas, later meetings might be to evaluate design details (e.g. icons etc.)

Usability Testing

Usability testing involves measuring typical user performance on a given task

User performance is typically measured in terms of the number of errors made and time to complete the task

As users perform tasks their interactions are recorded for later analysis

Users are observed by taking notes, using audio/video recordings and using monitoring software

Usability testing usually takes place in a lab where the evaluator is in control

It is important for the recording process that the user works on the task that the evaluator has chosen

Field Studies

Field studies are conducted in the natural settings of the users

The aim is to find out what users do with a product when not given a specific task to complete

Field studies may be conducted by observing users and recording what happens

The evaluator is an outsider to the user activity and collects data through qualitative/quantitative techniques

Field studies may be conducted by taking part in an activity and exploring the details

The evaluator is an insider (like in ethnographic studies) and tries to explore the design issues from within

Predictive Evaluation

Predictive evaluation involves an expert applying their knowledge of typical users

Experts provide their own “user models” based on their experience of users across many projects

Experts in user evaluation can provide useful insights to predict potential issues

Some issues that users have are so common that experts can quickly identify them without the need for testing

Experts provide both experience of users and an ability to interpret heuristics

Design heuristics can be difficult to interpret so the experience of an expert can be very helpful

Evaluation Techniques

Observing Users

Unobtrusively recording user interactions

Asking Users

Gathering user opinion and feedback

Asking Experts

Getting specific feedback from usability expert

User Testing

Gathering numeric data about user interactions

Modelling Users

Using models to avoid expensive testing

Observing Users

Observation techniques rely on being able to record important aspects of interaction

Recordings include written notes, photographs, audio and video recordings, computer logs (keystrokes etc.)

Evaluators try to observe users without disturbing or disrupting their activity

If the observation of users significantly changes their usual behaviour then the data gathered may be invalid

Evaluators must carefully analyse the data gathered from observation

Often there is a great deal of data that must be analysed and it may need to be combined from different sources

Asking Users

Asking users what they think of a product is an obvious way of getting feedback

e.g. Does it do what they want? Do they like the look of it? Did they have any problems using it?

Interviews and questionnaires are the main ways of asking users for their opinion

Questions can be structured or unstructured depending on the purpose of the evaluation

Users may also be encouraged to provide feedback by writing to the designers

It is common for websites to invite visitors to provide feedback via e-mail or via web-based forms

Asking Experts

Expert reviews and evaluations of designs can provide very useful feedback

Experts coming from outside the design team can bring new insights and perspectives about the design problem

Expert evaluations can involve role-playing users and applying heuristic methods

Expert evaluations are often quicker and less costly than conducting a laboratory or field testing

Experts frequently suggest solutions to design problems

User Testing

User testing is the principle method used in usability testing

i.e. user testing is typically conducted in a laboratory setting with users trying to complete a given task

Numerical data is collected about user performance

e.g. number of errors, time taken, number of keystrokes, number of links followed in a website, etc.

Statistical measures can be derived based on the data collected

Statistical measures can determine the importance and significance of different pieces of the data

Modelling Users

Models of users can be constructed for specific aspects of a design

e.g. a model of how long it takes for someone to perform certain actions at a computer

User models can provide early feedback without the need for evaluating prototypes

e.g. user models can be used to predict how long it will take an average user to perform a set of required actions

Framework

Determine the goals of the evaluation

What purpose will the evaluation serve?

Explore the questions to be answered

What questions do you want to be able to answer?

Choose an evaluation paradigm/techniques

What can be achieved given the time/budget available?

Identify the practical issues involved

How many users? How long will the evaluation take?

Decide how to deal with ethical issues

e.g. respecting a person's right to privacy

Evaluate, interpret and present the data

How reliable is the data? Could the evaluation be biased?

What are the limitations of the evaluation?

Pilot Studies

Small-scale pilot studies are often conducted before committing to large evaluations

Large-scale user evaluations can be very costly so it is wise to check the evaluation method of a small scale

Pilot studies can identify problems with the evaluation approach at an early stage

Pilot studies may identify problems with the selection of users, the operation of equipment, the phrasing of questions and many other aspects of the process

Summary

What, where, why and when to evaluate

Evaluation paradigms and techniques

DECIDE: A framework to guide evaluation