Experimental evaluation

Stephen Brewster

Glasgow Interactive Systems Group
School of Computing Science
University of Glasgow

stephen@dcs.gla.ac.uk

Why do Formal Experimental Evaluation?

Good for determining a causal relationship between interface characteristic and user performance/behaviour

Method by which human factors knowledge is gained and then turned into guidelines

Testing a component of the design under controlled conditions, typically in a usability lab

What are the Problems with Informal Evaluations?

Think Aloud

- subjective/qualitative
- gives you an idea of where problems are

Formal experiment

- objective/quantitative
- tells you exactly where problems are

Formal Experimental Evaluation

Quantitative technique

Investigate the relationship between two variables by changing one of them and looking at the effect of that change on the other

Can be formative or summative

Goal of experimental evaluation

Variables

Independent variable (IV)

Example - Icon design

- size of icon 16x16, 32x32, 64x64 pixels

Dependent variable (DV)

Example

- recall rate of icon
- speed of recall

Confounding Variables

Control for environment

Control for participants

Control for experimenter effects

Use a usability lab?

A Typical Usability Lab

Experimenter can control whole environment

Experimental Hypotheses

A prediction of the outcome of the experiment

Should be presented in terms of IV and DV

Experiment will show if prediction is correct

This is done by disproving the *Null Hypothesis* (H_0)

Participants

Who should they be?

How many?

- must be large enough to be representative of the user population

Participant	*Mean Scores*
1	5
2	4
3	30
4	6
Mean	**11.25**

Participant	*Mean Scores*
1	5
2	4
3	30
4	6
5	5
6	4
7	6
8	4
9	5
10	4
11	4
12	5
Mean	**6.833333333**

Experimental Designs

Main types

- Between subjects
- Within subjects

At least 2 conditions (depends on levels of IV)

Control condition

Experimental condition(s)

Between Subjects Design (Independent Samples)

Compare one group against the other

Participants randomly allocated to one of the groups

Each group of participants performs one condition

No learning problems

Condition 1	***Condition 2***
Participant 1	Participant 2
Participant 3	Participant 4

Example: Effect of Training Type on Usability

Hypotheses

- Web based, multimedia training more effective than text-based training
- users should need to consult the manual fewer times after multimedia training

IV - type of training

Example: Effect of Training Type on Usability

DV

number of times manual consulted when using system

Participants

24 users, novices with our new system

randomly allocated to the two groups

Between subjects design

Condition 1: Text	***Condition 2: Multimedia***
Participant 1	Participant 13
Participant 2	Participant 14
Participant 3	Participant 15
..	..
Participant 12	Participant 24

Within Subjects Design (Repeated Measures)

Compare each participant against him/her self

Participants perform each condition

Condition 1	***Condition 2***
Participant 1	Participant 1
Participant 2	Participant 2
Participant 3	Participant 3
Participant 4	Participant 4

Within Subjects Design (Repeated Measures)

Less chance of variation between participants

Within groups design generally better for small studies

Problem: each person can take a long time to test

Order Effects

Problem of learning - Order effect

Example: Text versus Multimedia training

Example: Time taken to learn the meanings of different sized icons

- IV - icons size, 3 levels
- Condition 1: small icons
- Condition 2: medium icons
- Condition 3: large icons

By the time you get to Condition 3 you have had two chances to learn the icons

Counterbalancing

Example: Investigating the Use of Sound in Buttons

Hypotheses

- time to recover from errors reduced with sound
- time to complete tasks reduced with sound

IV - style of buttons

- two levels:
- sound
- no sound (control)

Example: Investigating the Use of Sound in Buttons

DV

- time to recover from errors
- time to complete tasks

Participants

- 14 computer users

Within subjects design

Latin Square

Used to show the design of the experiment

	Condition 1	Condition 2
Group 1	Audio: Train and test	Control: Train and test
Group 2	Control: Train and test	Audio: Train and test

Stages in Preparing an Experiment

1. Choose hypothesis to test

2. Choose IV and work out levels, choose DV

3. Work out data to be collected

4. Choose experimental design

Stages in Preparing an Experiment

5. Choose tasks

6. Write experiment software

7. Recruit participants

8. Run the test

9. Analyse the results

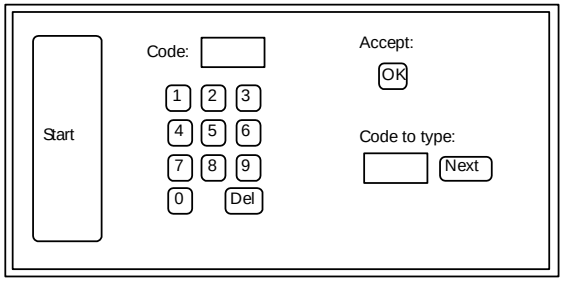
Experimental Tasks

Should be constrained to test just the thing you are interested in

Need to collect enough data to analyse

Example: Use of sound in buttons

- each condition 20 minutes
- as many buttons presses as possible in the time



Experimental Tasks

How to time condition

- you with a stopwatch
- timed within program

Data

- time when button pressed/released
- time taken to recover from an error
- save to a file

Running the Experiment

- Pilot the experiment
  - Testing a prototype of the expt
  - Fix it if you find problems
- Make sure each experimental run is as similar as possible to all of the others
- Put participants at their ease
- Get informed consent
- Give any training necessary
- Run the experiment and collect data on each participant

After the Experiment

After a formal experiment good to debrief participants

Analyse experimental results

- May be simple graphs/bar charts of results
- May detailed statistical analysis

Ecological Validity

Criticism of experimental approach is that it is artificial

When controlling variables can simplify things too much

Doing a controlled experiment in a lab may miss out on an important factor from the workplace

You must make sure that results of your experiment are general and can apply to the real world

How to Ensure Validity

Use externally valid tasks

- Come from interview or observation

Use sample of final user population

Ethics

Student projects and assessed exercises will sometimes involve other participants

Potential risks exist in such involvement

- Due to use of unusual or unfamiliar hardware
- Due to the environment where the task occurs
- Due to the nature of the participants
- Due to privacy or legal concerns

Ethics (2)

We have a responsibility to ensure the safety and well-being of those involved in, or affected by, our work

Ethics approval procedures ensure:

- Health and safety of subjects in projects
- Protection of investigators conducting projects
- That any relevant legal obligations are met

Ethics (3)

You must not cause your participants distress

- invading privacy, physical abuse, unpleasant emotions, etc.

Results from participants must be kept confidential

- you can the individual his/her results but you should not tell others, unless he/she agrees
- in reports you write do not refer to participants in any way that they could be identified

See the ethics checklist

www.dcs.gla.ac.uk/ethics

Ethics Checklist Form

A checklist of ethical issues to consider when doing projects that involve human subjects

www.dcs.gla.ac.uk/ethics/projects-form.pdf

Contains a list of 12 points to consider

Does your project conform to *all* points?

- Yes – sign the form, and include it with your project/ coursework submission
- No – you *must* request ethical approval form *before* proceeding

Resources

Test your understanding

- what is the main difference between informal and formal evaluation?
- define independent and dependent variables
- what is an hypothesis?
- why is a within-groups design more efficient than a between groups?
- what is ecological validity?

Resources

Robson, C. Experiment, design and statistics in psychology, 3rd Ed, Penguin Press

Dix, Finlay, Abowd & Beale, Human-Computer Interaction, Prentice Hall

Coolican, H. Research methods and statistics in psychology, Hodder & Stoughton

Heiman, G., Research methods in psychology, Houghton Mifflin

Many other books in the library on experimental design and statistics - some are pretty hard to understand!