

# Experimental Evaluation

By Richard Neves

## WHAT IS IT?

The purpose of an experimental evaluation is to concentrate on a single design-related issue. The experiment is controlled to test a specific hypothesis, e.g. it is quicker to speak rather than type in a Google command.

Through testing you measure an aspect of user behaviour to provide empirical evidence. The experimental conditions only differ in values of certain controlled variables; this is to suggest that any statistically significant changes in user behaviour are attributed to the different condition. It is important when designing an experiment that you are clear on what will change, the condition (independent variable), and what is constant. It is the resulting dependent variables that we measure in the experiment, e.g. time on task, number of errors, user rating on usefulness.

## WHY DO YOU NEED IT?

This approach is quite academic but does provide some benefits:

- To provide evidence that a design does/does not improve the user experience or efficiency.
- The alternative designs provide the independent variable you can compare based on time to complete the task or number of errors observed.

## WHEN DO YOU NEED IT?

This method would be performed early on, but when testing alternative designs it can be incorporated into the user testing phase.

## HOW DO YOU DO IT?

### Experiment Design

#### Between Groups

Different participants are used for each condition, although the groups should still be similar (i.e. same system experience, gender, age) – this is more costly because of the need for twice as many participants.

#### Within Groups

Same participants are used for each condition, a less costly approach, but based on the first condition, the participant may perform better on the second.

### Data Analysis

It is probably best to test the null hypothesis, i.e. that there is no difference between the conditions.

1. Calculate the means and standard deviation for each condition's dependent measures.
  2. Perform statistical analysis (t-test) for statistical or suggested statistical significance (for small participant numbers) between the conditions.
- Alternatively, average the dependent measures for each condition, calculate the standard deviation and make an informal comparison.

## EXAMPLE

---

|                             |  |
|-----------------------------|--|
| <b>Hypothesis</b>           | Search time decreases as the number of items on a menu decreases |
| <b>Independent variable</b> | Number of menu items   |
| <b>Dependent variable</b>   | Time to search   |
| <b>Conditions</b>           | Different levels: 3, 5, 7 items                                  |

---

Dix et al (2004) *Human-Computer Interaction*