Chapter 12. Testing in IT research

Table of Contents

Conducting tests in IT

Testing software and hardware is an important part of most IT research projects. It enables new ideas to be validated, alternatives to be compared against each other, problems to be identified and lessons to be learned for unanticipated behaviour or system properties.

The goals/questions tested should not be too broad (e.g. Is the system easy to learn? or is this algorithm faster?) specific issues should be addressed and then the test scenario can be easily formulated accordingly (eg. Are users able to discover $< X >$ in less than $< Y >$ minutes? or under conditions $< C >$, does algorithm $< A >$ complete task $< T >$ faster than algorithm $< B >$?)

What to measure

The measurements for evaluating the success of a software or hardware design will depend on the system and research question or goal. Speed, error rate, disk utilization, accuracy, scalability, reliability, usability and resource requirements are typical candidates for evaluation.

User testing is the most important technique for usability analysis. This is a type of experiment in which users are set specific tasks, and then their times are measured and errors they make are recorded. Other notes of interest may also be taken, possibly with the aid of observation, questionnaires or follow-up interviews noting for example the manner in which users perform a task when there are several possibilities.

It is important for user testing to measure performance, error rates and user satisfaction, since a user-friendly system is one which achieves a good balance between these three key factors. Data collected during user testing generally includes time to complete a task, number and types of errors made, number of errors made per unit time, number of times online help/documentation is used, percentage of users able to perform a task successfully.

Testing as scientific experiment

An experiment involves changing one variable and measuring its effect on another variable or property, while attempting to control all other variables/influences so they remain constant. In most cases, systems testing is performed in artificial laboratory situations, where system parameters, data and usage characteristics can all be controlled by the researcher. Software/hardware testing in the field tends to be confined to case studies; although field studies of safety-critical systems or of systems with major financial impact do also take place.

The human factor in usability analysis makes control more difficult, and researchers must take precautions to limit the impact of extraneous factors when designing and performing user testing. The following controls are typical of a usability study:

- The welcome, introduction and task questions are written out beforehand and are hence identical for all participants

- All subjects are given the same amount of time to explore the system freely at the outset (typically 5 to 10 minutes)

- All participants are given the tasks in the same sequence (usually from easiest to hardest, to build up confidence, and an easy one to end off with)

- Participants are limited to a maximum amount of time on a task, and are made to start the next task at that point

- Participants to not have contact with the outside world during the experiment

- All participants complete the same questionnaire after the experiment

- They are all asked their opinion of the system after the test is over

Differences between testing and scientific experiments

While usability/system testing is much like scientific experiment, it does not try to discover new knowledge, but rather to inform and improve system development. An experiment should be based on a theoretical foundation and targeted at solving a practical problem; its results should help to refine the theory and alleviate or solve the problem. The procedure should be documented and repeatable, and statistical analysis should be used to determine if results are significant in order to confirm a hypothesis.

In contrast, user testing is about finding problems and comparing alternatives in order to improve a product. Usability testing has relatively few participants (quick-and-dirty tests sometimes involve only one or two subjects, otherwise 6 to 12 users is the sample size recommended [Dumas and Redish 99]). The statistical analysis of usability test data also tends to be simpler than for other scientific experiments usually nothing more than calculating minimum, maximum, mean and standard deviation.

References: J.S. Dumas & J.C. Redish. A Practical Guide to Usability Testing (revised ed). Intellect:1999.