

i213 User Interface Design and Development

Professor Tapan Parikh
School of Information, UC Berkeley

USABILITY TESTING

Test interfaces with real users!

Basic process:

- Set a goal - what do you want to learn?
- Design some representative tasks
- Identify a set of likely users
- Observe the users performing the tasks
- Analyze the resulting data



FORMATIVE vs SUMMATIVE EVALUATION

Formative evaluation - Discover usability problems as part of an iterative design process. Goal is to uncover as many problems as possible.

Summative evaluation - Assess the usability of a prototype, or compare alternatives. Goal is a reliable, statistically valid comparison.

I NEED YOUR HONEST
FEEDBACK ON OUR
NEW WEBSITE DESIGN.



Dilbert.com DilbertCartoonist@gmail.com

THE LAYOUT LOOKS
LIKE A PSYCHOPATH'S
PHOTO WALL. THE
COLORS REMIND ME
OF TOE FUNGUS AND
DESPAIR.



© 2010 Scott Adams, Inc./Dist. by UFS, Inc.

I'LL
SAY,
"NEEDS
WORK."

IT FEELS
LIKE SATAN
IS LICKING
MY BRAIN!



PERFORMANCE MEASUREMENT

Implies testing a user interface to obtain statistics about performance

Most useful for summative evaluation

Can be done to either:

- Compare variants or alternatives
- Decide whether an interface meets pre-specified performance requirements

PERFORMANCE METRICS

Time to complete the task

Number of tasks completed

Number of errors

Number of commands / features used

Number of commands / features not used

Frequency of accessing help

Frequency of help being useful

Number of positive user comments

Number of negative user comments

Proportion of users preferring this system

etc...

DESIGNING AN EXPERIMENT

Independent variables (Attributes) - the factors that you want to study

Dependent variables (Measurements) - the outcomes that you want to measure

Levels - Acceptable values for measurements

Replication - How often you repeat the measurement, in how many conditions, with how many users, etc.

NUMBER OF VARIABLES

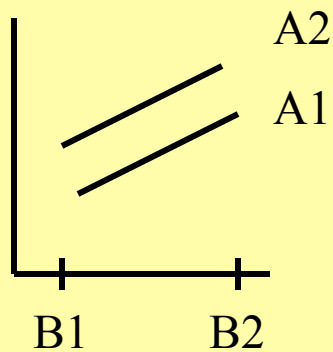
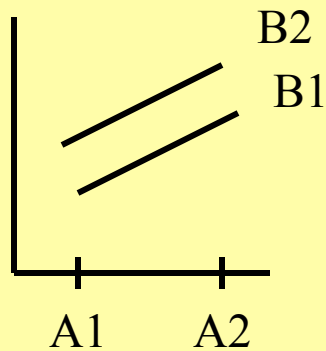
Varying one attribute (ex. color) is simple - consider each alternative for that attribute separately

Varying several attributes (ex. color, font and font size) can be more challenging:

- Interaction between attributes
- Blowup in the number of conditions

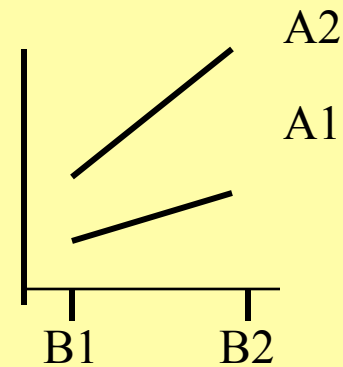
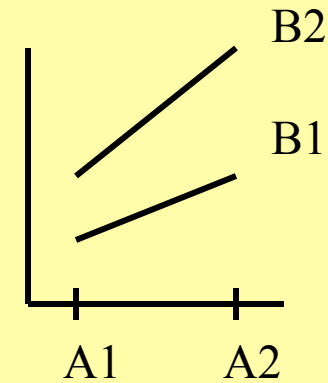
A and B do not interact

	A1	A2
B1	3	5
B2	6	8



A and B may interact

	A1	A2
B1	3	5
B2	6	12



DEALING WITH ATTRIBUTES

Conduct pilot tests to see which are important and interact

Take these attributes, and organize them in a Latin square, addressing ordering and making sure all variations are tested

Note: each user may only see a subset of the variations, and only some orderings may be considered

	T1	T2	T3	T4
6	G	G+	A	A+
6	G+	A	A+	G
6	A	A+	G	G+
6	A+	G	G+	A

BETWEEN vs. WITHIN SUBJECTS

When comparing two interfaces

Between-Subjects: Distinct user groups use each variation

- Need large number of users to avoid bias in one sample vs. the other
- Random vs. matched assignment

Within-Subjects: Same users use both variations

- Can lead to learning effects
- Solution is to *counter-balance* the study - each group uses one interface first

CONDUCTING A PILOT TEST

Before unleashing your system and your testing scheme on unwitting users, it helps to pilot test your study

Iron out any kinks - either in your software, or your testing setup

A pilot test can be conducted with design team members and other readily available people (at least one of them should be a potential user)

SELECTING USERS

“Should be as representative as possible of the intended users”

If testing with a small number of users, avoid outlier groups

If testing with a larger number of users, aim for coverage of all
“personas”

Include novices, probably experts too

It helps if users are already familiar with hardware (if its not part of
your design)

SOURCES OF TEST USERS

Early adopters

Students

Retirees

Paid volunteers

... *Be creative!*



HOW TO TREAT USERS

Train them if you will assume some basic skills (ex. using a mouse)

Do not blame or laugh at the user

Make it clear that the system is being tested, not the user

Make the first task easy

Inform users that they can quit anytime

After the test, thank the user

HUMAN SUBJECTS

In many universities and research organizations, UI testing is treated with similar care as medical testing

Requires filling out and submitting a Human Subjects approval form to the appropriate agency

Important considerations include maintaining the anonymity of test users, and obtaining informed consent

STATEMENT OF INFORMED CONSENT

If you volunteer to participate in this study, you will be asked to perform some tasks related to XXX, and to answer some questions. Your interactions with the computer may also be digitally recorded on video, audio and/or with photographs.

This research poses no risks to you other than those normally encountered in daily life. All of the information from your session will be kept anonymous. We will not name you if and when we discuss your behavior in our assignments, and any potential research publications. After the research is completed, we may save the anonymous notes for future use by ourselves or others.

Your participation in this research is voluntary, and you are free to refuse to participate or quit the experiment at any time. Whether or not you chose to participate will have no bearing in relation to your standing in any department of UC Berkeley. If you have questions about the research, you may contact X at Y, or by electronic mail at Z. You may keep a copy of this form for reference.

If you accept these terms, please write your initials and the date here:

INITIALS _____

DATE _____

HELPING USERS

Decide in advance how much help you will provide (depending on whether you plan to measure performance)

For the most part you should allow users to figure things out on their own, so tell them in advance that you will not be able to help during the test

If user gets stuck and you aren't measuring, give a few hints to get them going again

Terminate the test if the user is unhappy and not able to do anything

User can always voluntarily end the test

CONCERNS WITH USERS

People get tired!

People get bored!

People can get frustrated!

People can get distracted!

People learn how to do things!

All of these can be exacerbated in a Within-Subjects test

DESIGNING TEST CASES

Should be representative of real use cases

Small enough to be completed in finite time, but not so small that they are trivial

Should be given to the user in writing, to ensure consistency and a ready reference

(Don't explain how to do it though!)

Provide tasks one at a time to avoid intimidating the user

Relate the tasks to some kind of overall scenario for continuity

EXAMPLE TASK DESCRIPTION

Motivating Scenario: “Use your mobile phone for accessing and editing contact information.”

Tasks:

1. Find the contacts list in the phone.
2. View the contact information for John Smith.
3. Change John Smith’s number to “555-6666”.
4. ...

STAGES OF A USABILITY TEST

Preparation

Introduction

Observation

Debriefing

PREPARATION

Choose a location that is quiet, interruption-free, and has all the equipment that you need

Print out task descriptions, instructions, test materials and/or questionnaires

Install the software, and make sure it is in the “start” position for the test

Make sure everything is ready before the user shows up

INTRODUCTION

Explain the purpose of the test

Ask user to fill out the Informed Consent form, and any pre-test surveys (including demographics)

Ensure the user that their results will be kept confidential, and that they can stop at any time

Introduce test procedure and provide written instructions for first task

Ask the user if they have any questions

CONDUCTING THE TEST

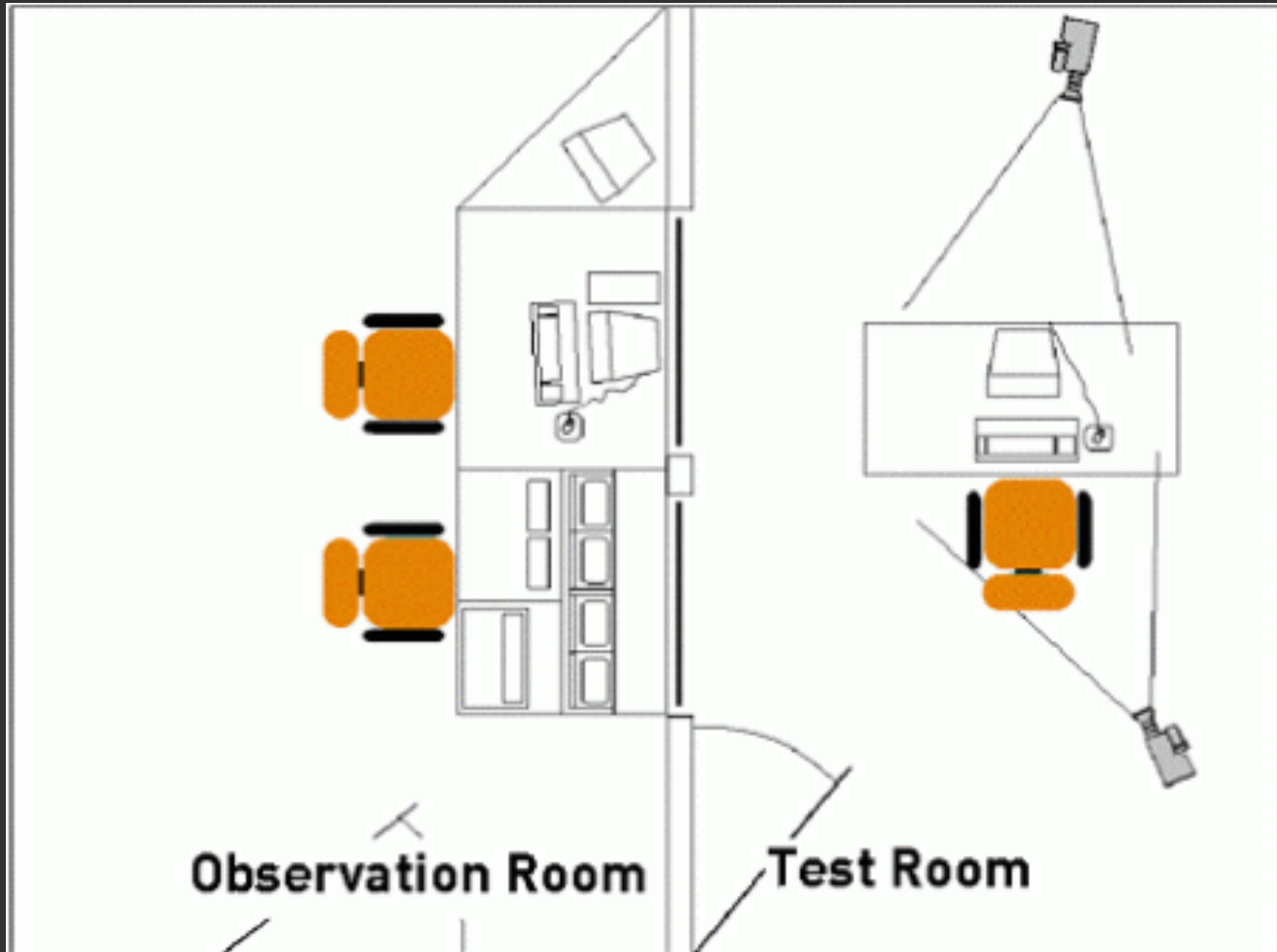
Assign one person as the primary experimenter, who provides instructions and communicates with the user

Experimenter should avoid helping the user too much, while maintaining a positive attitude

No help can be given when performance is being measured

Make sure to take notes and collect data!

Usability Labs



CONDUCTING THE TEST

Administer subjective satisfaction questionnaires, often using Likert scale

- Rate your response to this statement on a scale of 1-5, where 1 means you disagree completely, and 5 means you agree completely

“I really liked this user interface!”

Ask user for any comments or clarification about interesting episodes

Answer any remaining user questions

Disclose any deception used in the test

Label data and write up your observations

Semantic differentials

The task that we asked you to perform was...

	very	reasonably	neither-nor	reasonably	very	
clear	1	2	3	4	5	unclear
complex	1	2	3	4	5	simple
familiar	1	2	3	4	5	unfamiliar

The search process you have just performed was...

	very	reasonably	neither-nor	reasonably	very	
stressful	1	2	3	4	5	relaxing
interesting	1	2	3	4	5	boring
tiring	1	2	3	4	5	restful
easy	1	2	3	4	5	difficult

The summaries/abstracts presented by the system were...

	very	reasonably	neither-nor	reasonably	very	
irrelevant	1	2	3	4	5	relevant
unimportant	1	2	3	4	5	important
not useful	1	2	3	4	5	useful
incomplete	1	2	3	4	5	complete

VALIDITY

Validity can be impacted by setting up the wrong experiment

- Wrong users
- Wrong tasks
- Wrong setting
- Wrong measurements
- Confounding effects

Take care in your experimental design about what you are testing, with whom, and where

RELIABILITY

Reliability of results can be impacted by variation amongst users

- Include more and more diverse users
- Use standard statistical methods to estimate variance and significance

Confidence intervals are used for studies of one system

Pairwise T-test is often used for comparing two systems

DESIGNERS AS EVALUATORS

Usually system designers are not the best evaluators

Potential for helping users too much, or explaining away usability problems

Evaluator should be trained in the evaluation method, and also be an expert in the system being tested

Can be a team of a designer and an evaluator, who handles user relations

TESTING YOUR PROJECTS

- 1) What are important metrics to assess?
- 2) What are the alternatives to consider, including the current de facto?

For next time

Keep working on functional prototype

Guest lecture on crowd-sourcing