User Studies

Ben Bederson
CMSC434 - Intro to HCI
Spring 2011

Slides (slightly) modified from Scott Klemmer

Relative Merits: Online vs. In-lab?

Scale
Control (confounding variables)
Realism

Ethical Considerations

Sometimes tests can be distressing: users have left in tears. You have a responsibility to alleviate: make voluntary with informed consent avoid pressure to participate let them know they can stop at any time stress that you are testing the system, not them make collected data as anonymous as possible

You are testing the site (not the users)

Articulate the Goals of Your Study

Scope
 What are you testing?
Purpose
 What concerns, questions, and goals is the test focusing on?
Schedule and location
 When and where will the test take place?
Participants
 How many users of what types will you recruit?
Scenarios
 What will participants do with the product in this round of testing?

Create Concrete Tasks

A good plan for usability testing gives the participants:
a goal/task (what to do or what question to find the answer for)
data, if needed, that a real user would have when going to the site to do that task
You can give the scenario as just the statement of the goal/task or you can elaborate it a little with a very short story that adds motivation to get to the goal.

Plan Out the Steps

Questions

What will you ask at the beginning and end of the session?

Data to be collected

What will you count?

Set up

What system will you use for testing? Will you be videotaping and/or audiotaping? Will you be using a specific technology to capture data?

Roles

Who will do what in the usability test?

Experimental Details

Order of tasks
choose one simple order (simple -> complex)
unless doing within-subjects experiment
Training
depends on how real system will be used
What if someone doesn't finish?
Pilot study
helps you fix problems with the study
do 2, first with colleagues, then with real users

Options for capturing results

Think aloud
Nothing critical incidents
Video recording
Screen recording
Decide whether to interrupt or not

The “Thinking Aloud” Method

- Need to know what users are thinking, not just what they are doing
- Ask users to talk while performing tasks
  - tell us what they are thinking
  - tell us what they are trying to do
  - tell us questions that arise as they work
  - tell us things they read
- Make a recording or take good notes
  - make sure you can tell what they were doing

Thinking Aloud (cont.)

- Prompt the user to keep talking
  - "tell me what you are thinking"
- Only help on things you have pre-decided
  - keep track of anything you do give help on
- Recording
  - use a digital watch/clock
  - take notes, plus if possible
    - record audio & video (or even event logs)
- Will thinking aloud give the right answers?
  - not always
  - if you ask a question, people will always give an answer, even it is has nothing to do with facts
  - try to avoid specific questions

Recruiting participants

The participants must be like the people who will use your site.

You might end up using a screening questionnaire

... plan on a cost associated with finding the people ... you may still need to plan on incentives to get participants to participate ...

Before “real” users...

Run a pilot to get out the kinks

Setting up the test sessions

Make sure you have everything you need

the prototype you are going to test

the computer set up for the participant with the monitor, resolution, and connection speed that you indicated in the test plan

note-taking forms on paper or set up on a computer

consent forms for participants to sign and a pen in case the participant does not bring one

questionnaires, if you are using any

the participant's copy of the scenarios

cameras, microphones, or other recording equipment if you are using any

folders to keep each person's paperwork in if you are using paper

Greeting Participants

The facilitator:
welcomes the participant and introduces anyone else who is in the room
invites the participant to sit in front of the computer where the participant will be working
explains the general goal of the session—to have the participant try out a Web site (or whatever the product is that is being tested)
asks participant profile questions and has the participant sign the release form
explains thinking aloud (and may demonstrate it and have the participant do a think aloud exercise)
asks if the participant has any questions before starting and answers any that will not give away what you want to learn from the participant
tells the participant where to start

Instructions to Participants

Describe the purpose of the evaluation
“I’m testing the interface; I’m not testing you”
Tell them they can quit at any time
Demonstrate the equipment
Explain how to think aloud
Explain that you will not provide help
Describe the task
give written instructions, one task at a time

Conducting the usability test

- The facilitator:
  - welcomes the participant and introduces anyone else who is in the room
  - invites the participant to sit in front of the computer where the participant will be working
  - explains the general goal of the session—to have the participant try out a Web site (or whatever the product is that is being tested)
  - asks participant profile questions and has the participant sign the release form
  - explains thinking aloud (and may demonstrate it and have the participant do a think aloud exercise)
  - asks if the participant has any questions before starting and answers any that will not give away what you want to learn from the participant
  - tells the participant where to start
- The participant starts to work with the Web site (or other product).
  - The participant works on the scenario while thinking aloud. The note-takers take notes.
  - The session continues from scenario to scenario until the participant has done (or tried) them all or the time allotted has elapsed.
- The facilitator asks the end-of-session questions
  - thanks the participant, giving the participant the agreed-on incentive, and escorts the participant out.

Source: Usability.gov

Collecting Data

- process data
  - observations of what users are doing & thinking
- bottom-line data
  - summary of what happened (time, errors, success)
  - i.e., the dependent variables

Debriefing

Debriefing test users
often don’t remember, so demonstrate or show video segments
ask for comments on specific features
show them screen (online or on paper)

Analyzing the results

- Quantitative data, which might include:
  - success rates
  - time to complete tasks
  - pages visited
  - error rates
  - ratings on a satisfaction questionnaire
- Qualitative data, which might include:
  - notes of your observations about the pathways participants took
  - notes about problems participants had (critical incidents)
  - notes of what participants said as they worked
  - participants' answers to open-ended questions

Source: Usability.gov

Summarize the Results

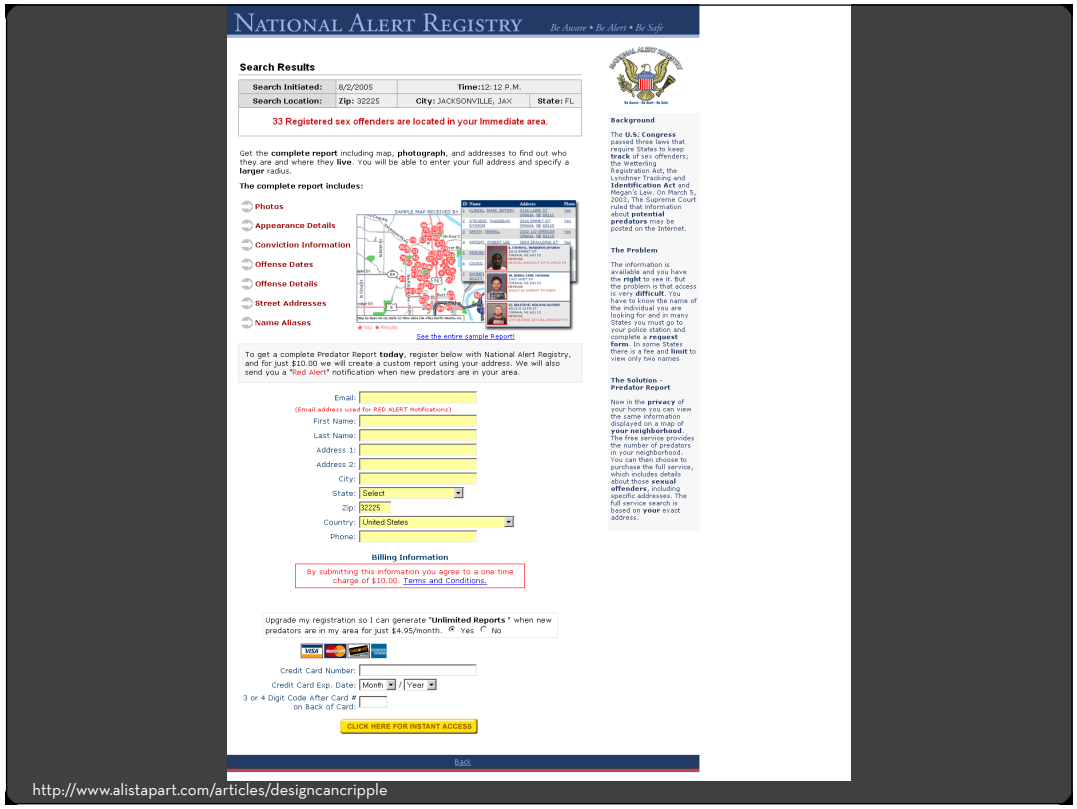
Summarize the data
make a list of all critical incidents
positive & negative
include references back to original data
try to judge why each difficulty occurred
What does data tell you?
UI work the way you thought it would?
users take approaches you expected?
something missing?
Update task analysis & rethink design
rate severity & ease of fixing critical incidents
fix both severe problems & make the easy fixes

Reporting the Results

Report what you did & what happened Images & graphs help people get it! Video clips can be quite convincing

Controlled Web Experiments

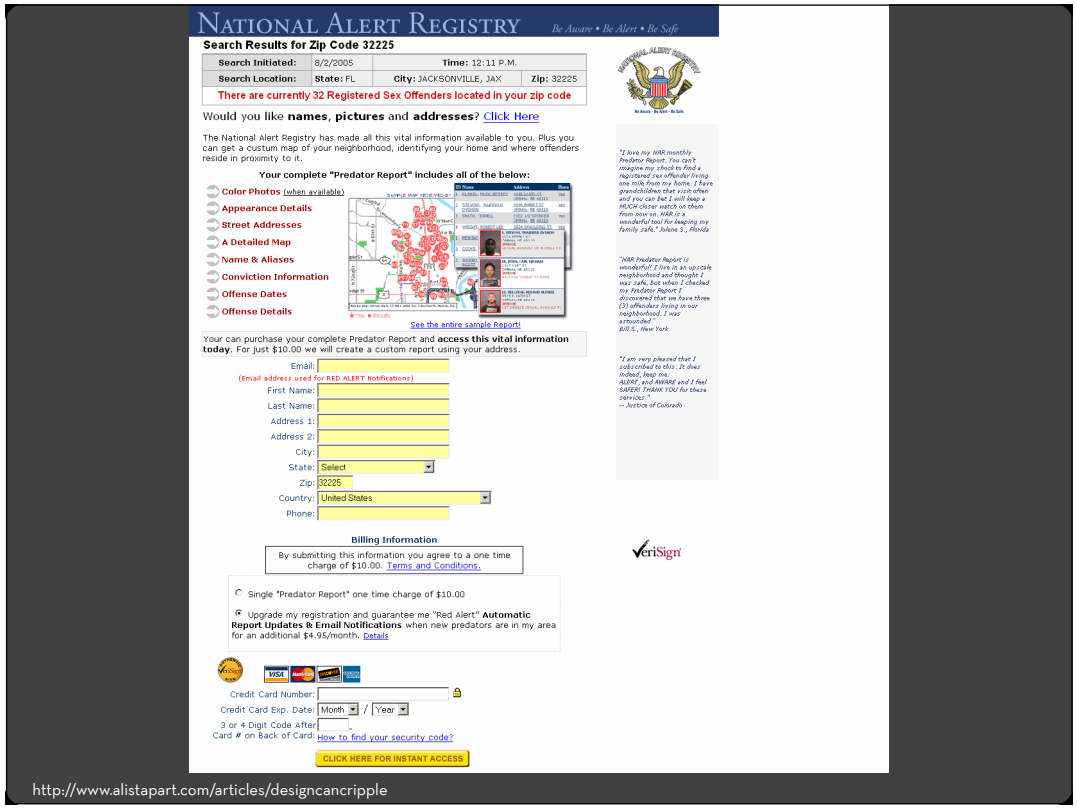
Many names; same idea
A/B tests
Randomized experiment
Controlled experiment
Split testing
Randomly split traffic between two versions
A/Control: usually current live version
B/Treatment: new idea
Collect metrics, analyze



http://www.alistapart.com/articles/designcancripple

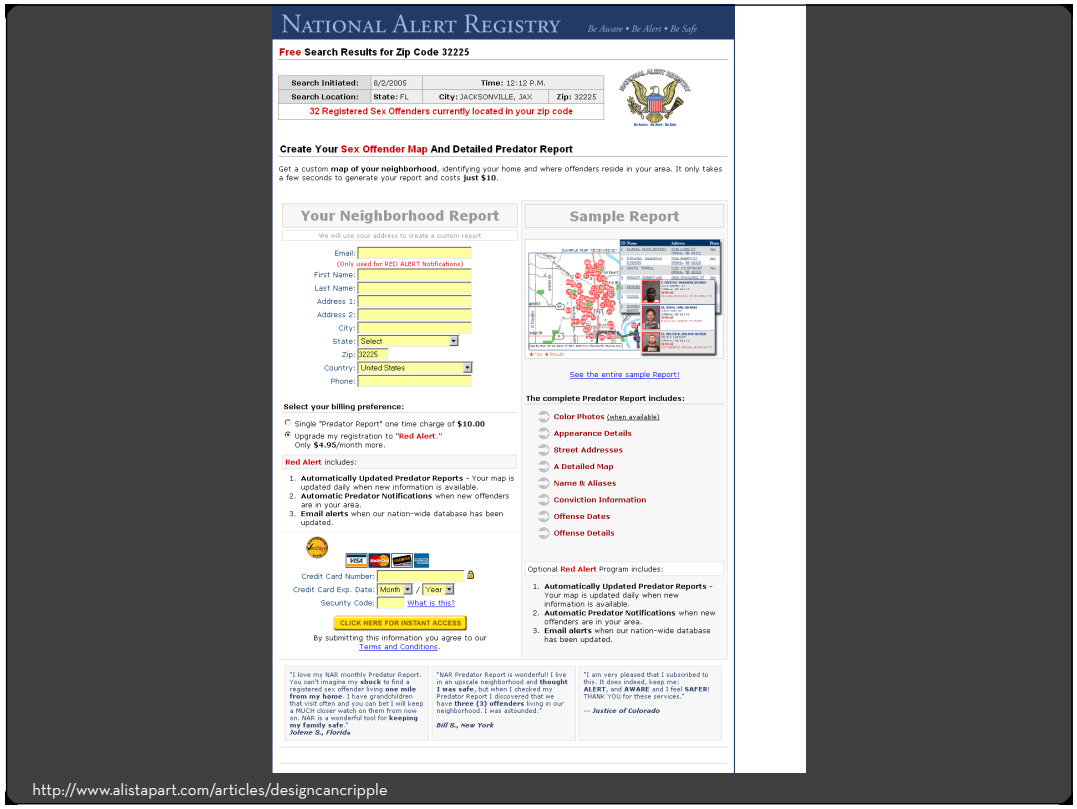
However compelling the message, however great the copy, however strong the sales argument... the way a page is designed will have a dramatic impact on conversion rates, for better or for worse. Here are three versions of the same offer pageI know, they won't win any design awards. They weren't intended to. But they are functional and familiar. A reader going to any one of these pages will be able to quickly figure out what the message is, and what they are being asked to do.

Version A is the original.



http://www.alistapart.com/articles/designcancripple

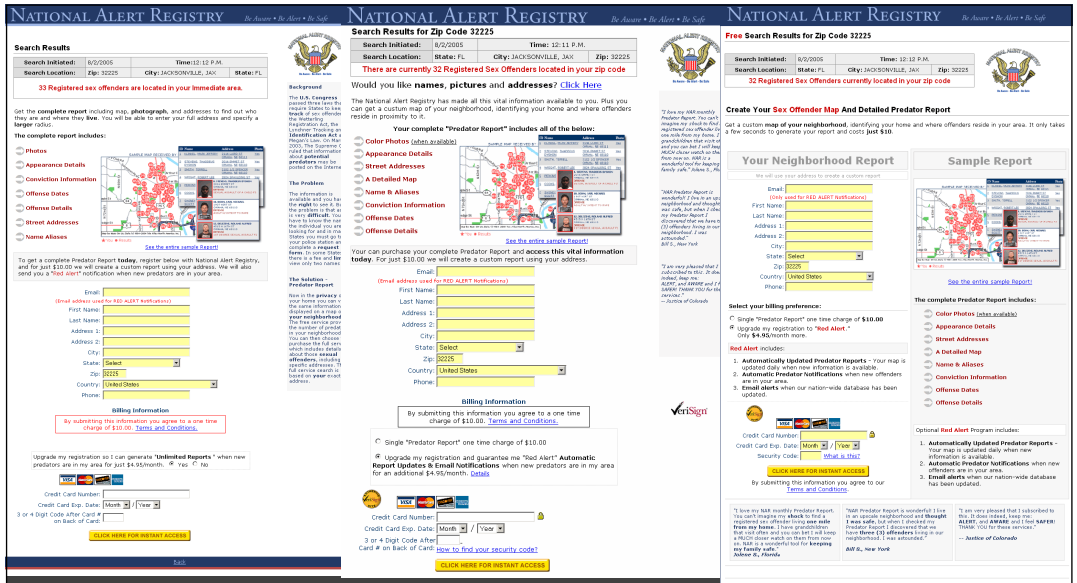
Version B follows the same basic layout, but we made some minor copy changes.



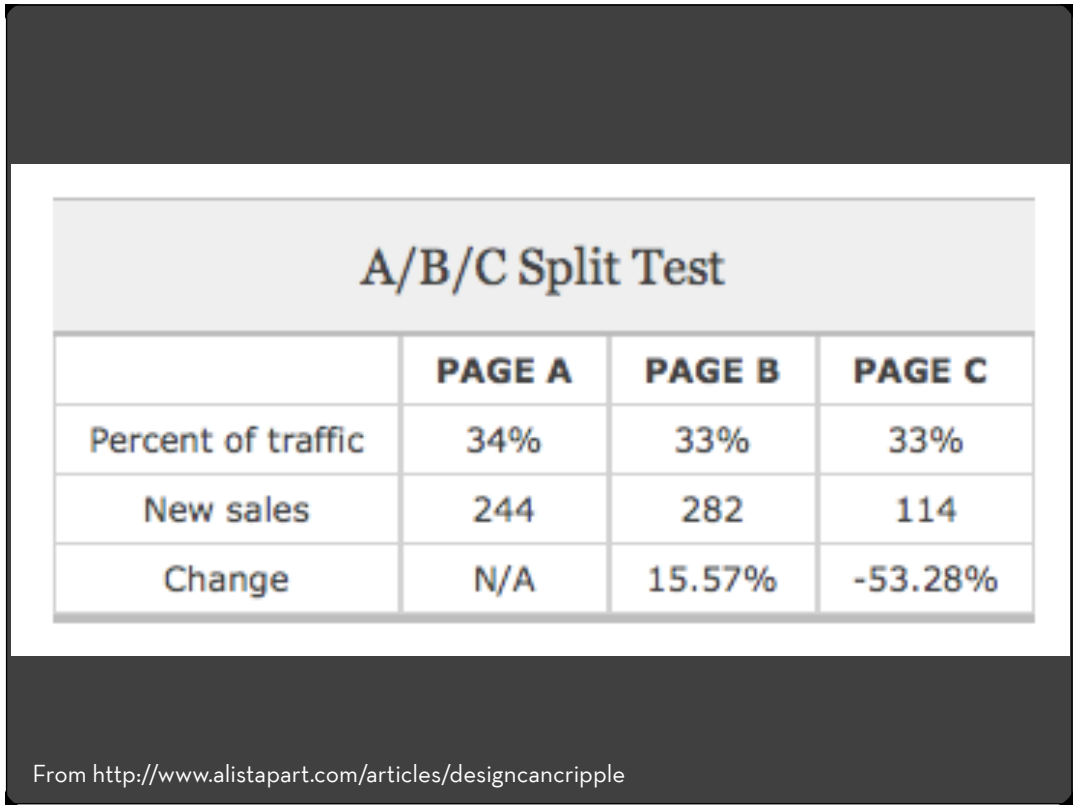
http://www.alistapart.com/articles/designcancripple

In version C, we changed from a one-column format to two-column format. We wanted to test the impact of bringing more of the page content onto the first screen. Be honest with yourself and decide now whether B or C beat A, and by what percentage. I imagine you have some way of measuring the success of your site. Maybe it's about sales. Maybe it's based on readership. But one way or another, your site has a purpose. But I don't think most designers truly understand the effect their design choices can have on achieving that purpose. And yes, I'm sure you do some usability testing. And that likely gives you some broad, if sometimes confusing insights into what's working and what isn't. But do you test different page designs?

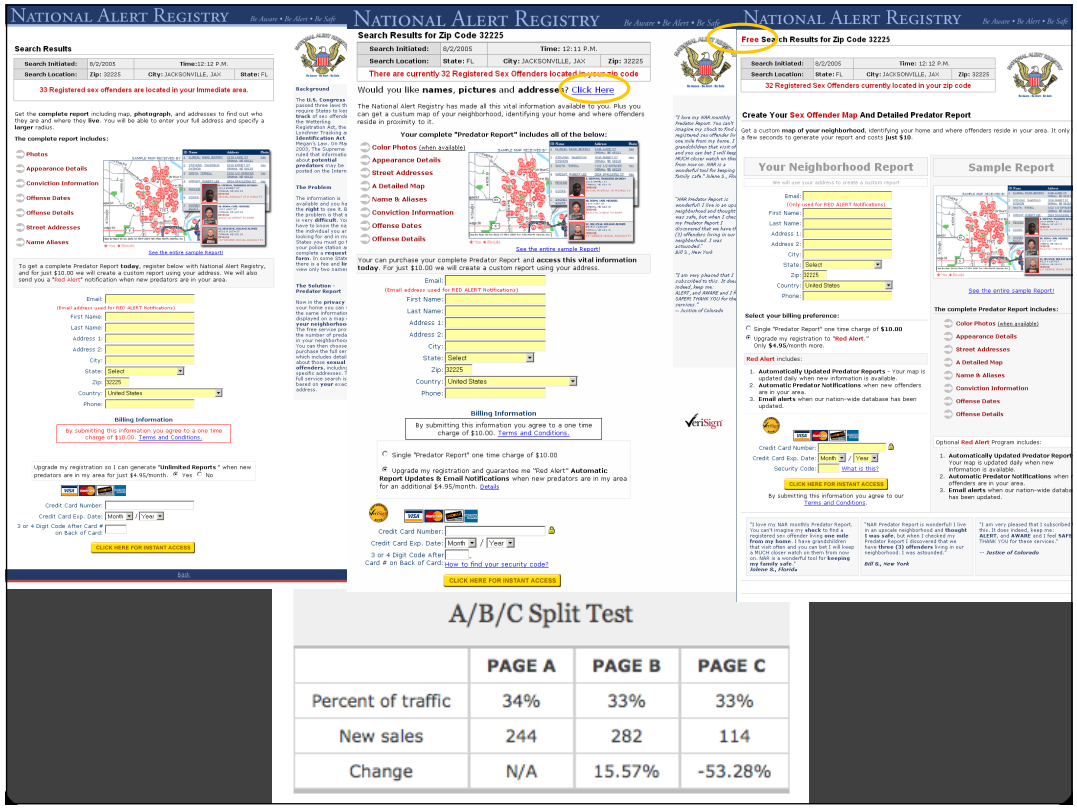
By testing, I don't mean asking a few folks around the office; I mean doing a live test that demonstrates—with hard figures—what site visitors actually do. Testing like that is a beautiful thing. There is no space for fancy arguments. An expert's credentials and opinions mean squat. When you serve alternative versions, one after the other, and measure reader actions, you get the real deal. You get what is. But if you are serious about achieving your site's purpose, and if testing can show you which version of a page does best, then where is the argument not to test?



From http://www.alistapart.com/articles/designcancripple



“That’s an astonishing reduction in sales and revenues, resulting from a design change that was intended to improve the performance of the page.
Now, just pause for a moment and think of all the design choices you have made over the last year, and the reasons why you made them. And think about the huge impact those choices might have had on the performance of the sites you worked on.”



	PAGE A	PAGE B	PAGE C
Percent of traffic	34%	33%	33%
New sales	244	282	114
Change	N/A	15.57%	-53.28%

Why did B do better and C do so much worse? A few theories. One is the free. Maybe another is it's more confusing. It employs a "menu" pattern when what's needed is a "funnel" pattern. B has a call to action, "would you like…? click here" at the top.

Ways design makes a difference

- The position and color of the primary call to action
- Position on the page of testimonials, if used
- Whether linked elements are in text or as images
- The amount of “white space” on a page, giving the content space to “breathe”
- The position and prominence of the main heading
- The number of columns used on the page
- The number of visual elements competing for attention
- The age, sex and appearance of someone in a photo

From http://www.alistapart.com/articles/designcancripple