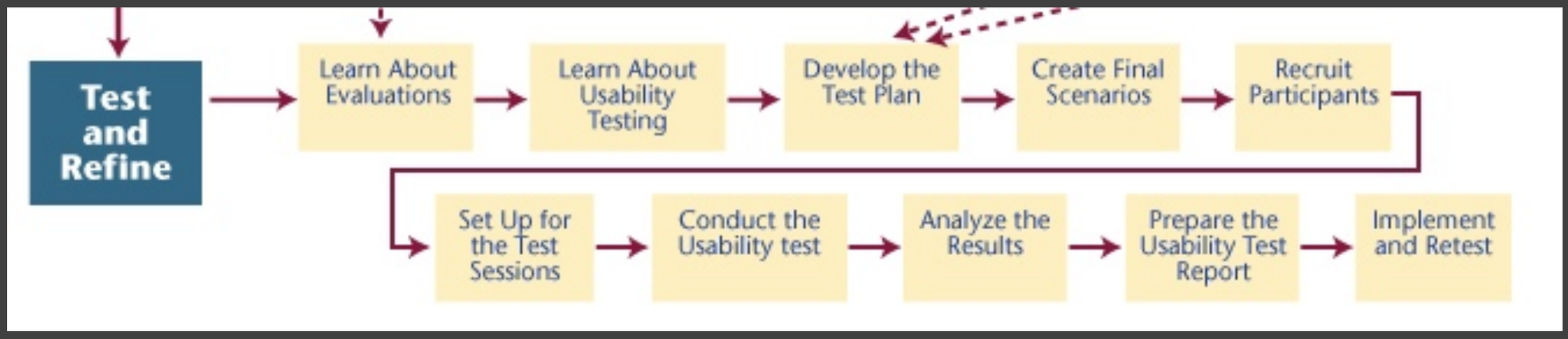
Conducting a User Study

Scott Klemmer

Autumn 2009

Step-by-step testing guide



Source: Usability.gov, http://www.usability.gov/process.html

Setting goals / developing the test plan

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?

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?

Source: Usability.gov

Creating a plan

- 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.

Source: Usability.gov

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 …

Source: Usability.gov

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
- Do a dry-run and a pilot test

Source: Usability.gov

Selecting Tasks

- Should reflect what real tasks will be like
- Tasks from analysis & design can be used
  - may need to shorten if
    - they take too long
    - require background that test user won't have
- Try not to train unless that will happen in real deployment
- Avoid bending tasks in direction of what your design best supports
- Don't choose tasks that are too fragmented
  - e.g., phone-in bank test

Summary: Before Starting

- You should know, and have written down
  - objective
  - description of system
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures
- Seems tedious, but writing this will help “debug” your test

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

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

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

Options for capturing results

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

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

Details (cont.)

- Keeping variability down
  - recruit test users with similar background
  - brief users to bring them to common level
  - perform the test the same way every time
    - don’t help some more than others (plan in advance)
  - make instructions clear
- 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)

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

Measuring Bottom-Line Usability

- Situations in which numbers are useful
  - time requirements for task completion
  - successful task completion
  - compare two designs on speed or # of errors
- Ease of measurement
  - time is easy to record
  - error or successful completion is harder
    - define in advance what these mean
- Do not combine with thinking-aloud. Why?
  - talking can affect speed & accuracy

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

Using the Test 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

Analyzing the Numbers

- Example: trying to get task time <=30 min.
  - test gives: 20, 15, 40, 90, 10, 5
  - mean (average) = 30
  - median (middle) = 17.5
  - looks good!
- Wrong answer, not certain of anything!
- Factors contributing to our uncertainty
  - small number of test users (n = 6)
  - results are very variable (standard deviation = 32)
    - std. dev. measures dispersal from the mean

Analyzing the Numbers (cont.)

- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that “typical” value is between 5 & 55

Analyzing the Numbers (cont.)

Web Usability Test Results						
Participant #		**Time (minutes)**				
1		20				
2		15				
3		40				
4		90				
5		10				
6		5				
	number of participants	**6**				
	mean	**30.0**				
	median	**17.5**				
	std dev	**31.8**				
	standard error of the mean	**= stddev / sqrt (#samples)**			**13.0**	
	typical values will be mean +/- 2*standard error			**--> 4 to 56!**		
	what is plausible? = confidence (alpha=5%, stddev, sample size)	**25.4**	**--> 95% confident between 5 & 56**			

Analyzing the Numbers (cont.)

- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that typical value is between 5 & 55
- Usability test data is quite variable
  - need lots to get good estimates of typical values
  - 4 times as many tests will only narrow range by 2x
    - breadth of range depends on sqrt of # of test users
- this is when online methods become useful
  - easy to test w/ large numbers of users

Reporting the Results

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