How to Add Usability Testing to Your Evaluation Toolbox

Christine Andrews Paulsen, Ph.D.

Concord Evaluation Group

cpaulsen@concordevaluation.com

Presented at AEA, 11/5/11, Anaheim, CA

Goals

- Develop an understanding of usability testing (UT) and why it's important

- Learn basics of developing testing script

- Learn fundamentals of conducting UT sessions

- Learn tips for analyzing data

- ...all before lunch!

Overview

- Many of the programs and initiatives that we evaluate today are technology-based.
- Interventions are delivered via...
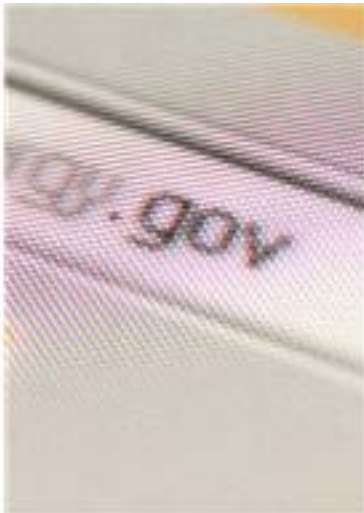
websites

mobile devices

software apps

Technology-based evaluations should include usability testing

- To properly evaluate such initiatives, the evaluator must consider the usability (user-friendliness and accessibility) of the technology components.
- "Usability refers to how well users can learn and use a product to achieve their goals and how satisfied they are with that process."

Why Usability Testing?

- UT methods enable us to gather important **formative** data on technological tools (can be iterative).

- TIP: **Problems addressed early are less expensive to fix.** (High ROI)

- UT methods also help us explain outcomes and impact during **summative** evaluation.

Usability Objectives

- UT helps you evaluate:
  - Information architecture, organization
  - Presentation
  - Navigation
  - Accessibility
  - Support features
  - Special functionality
  - Visual design

Examples

- Browser incompatibility
- Software that is not accessible to deaf students
- Pages are too long/content too dense to find anything
- Navigation is confusing, can't find your way
- Colors make it too hard to read

One-on-One Usability Testing

- This is only one type of usability research method – but, probably the most common.
- **Goal: To observe a "representative" sample of users attempt to complete tasks while you observe and record data.**
- Relies heavily on the "think aloud protocol" (Ericsson & Simon, 1993)

Test Setting

- Can be in-person or remote (online)
- What you need:
  - Testing script
  - Trained researcher
  - Participant
  - Evaluand (the tool/media to be tested)
  - Privacy and no interruptions
  - Recording devices (paper, digital recorder)
  - Optional: A way for observers to watch unobtrusively

Setting, continued

In-person

Remote

Participant

Researcher

One-way mirror

Observers

Participant

Researcher

Observers
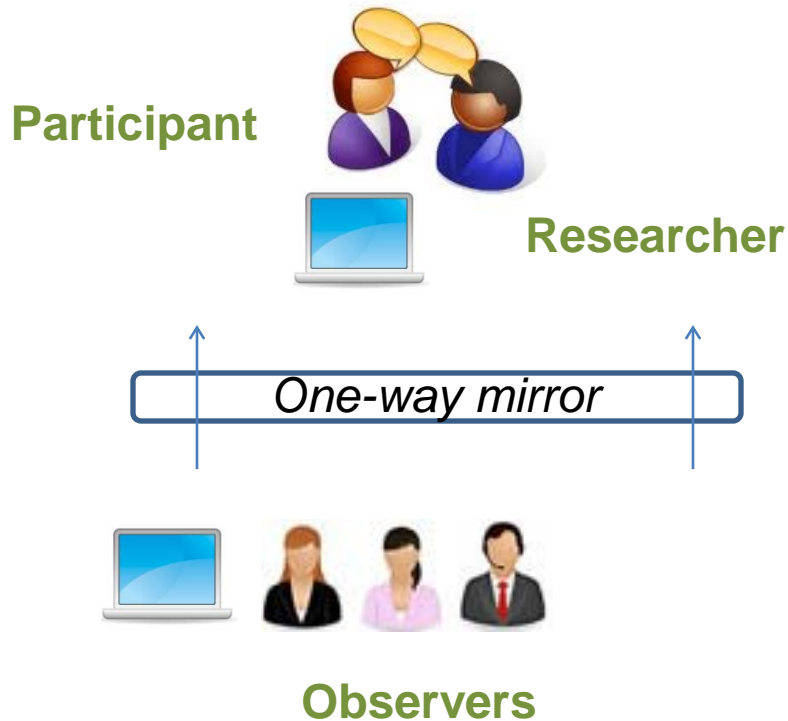
UT Test Plan (Script)

- The test plan documents **methods and procedures**, **metrics** to be captured, number and type of **participants** you are going to test, and what **scenarios** you will use. Elements include:
  - Scope (What is Being Tested)
  - Purpose & Objectives
  - Schedule & Location
  - Length of Sessions
  - Equipment
  - Number & Type of Participants
  - ***Scenarios & Task Completion Paths***
  - Metrics (Subjective & Objective)
  - Team Member Roles

TIP: Templates are available at Usability.gov

Scenarios

In collaboration with the client or organization being evaluated…

- Step 1: Make a list of the top 10 tasks users try to complete when using the technology.
  - Example: "Users come to the site to find reports on different topics related to education."
- Step 2: Develop a list of roughly 10 known issues or problems that could be tested to study whether the issues are usability problems.
  - Example: "Some users have reported that the pages are too long and difficult to read."
- Step 3: Write a set of 20 scenarios for pilot testing. Good scenarios are brief and easy to understand.
  - Example: "Using the search engine, please locate 5 articles on school safety."
  - Example: "Imagine you are developing a training on assessment, please use the website to locate a training template."
- Step 4: After pilot testing, eliminate or revise scenarios that were ineffective. Plan on using 10-15 scenarios for a typical UT.
- Step 5: You will need to record the task scenario completion paths so the moderator knows what to expect and how to determine successful completion of a task. Participants will not see the paths.

Your Turn

- Think about a technology that you are evaluating or one that you use often
- Think about some of the most common problems you know of
- Think about some of the most common tasks/reasons people use this technology
- Share some scenarios for discussion

Recruitment

In collaboration with the client or organization being evaluated…

- Step 1: Identify the target population of technology users.
- Step 2: Determine the sample size. During formative testing, 6-8 users *per homogenous subgroup* will typically uncover most usability problems. The sample size will increase if inferential statistics are needed.
- Step 3: Consider all recruitment costs include the cost associated with finding participants, incentives to get them to come (e.g., gifts or money), and in some cases travel/parking expenses.
- Step 4: Screen, enroll and schedule participants.

Facilitating the Test

- Step 1: Welcome the participant and obtain informed consent.
- Step 2: Explain the test session procedures.
- Step 3: Demonstrate the think aloud procedure (crucial).
- Step 4: The participant will begin by reading the task scenario and will work on completing it while they think aloud.
- Step 5: Researcher will take notes of the participant's behaviors, comments, errors and completion (success or failure).
- Step 6: The UT session continues until all task scenarios are completed or time allotted has elapsed.
- Step 7: The facilitator asks the end-of session subjective questions, thanks the participant, and gives the participant the agreed-on incentive (or makes arrangements to have it sent).

Tips

- Treat participants with respect and make them feel comfortable.
    - "We are evaluating the technology, not you."
- Remain neutral.
    - "We are interested in what *you* would do in this situation."
- Only offer help if the participant is frustrated.
    - Record data as "Completed with assistance."

Recording Data

- Non-verbal
- Verbalizations
- Paths
- Completed, Completed with Help, Not Completed
- Time

Demo

- Let's practice

Analyzing the Data

- Goal is to make recommendations about…
  - Information architecture, organization
  - Presentation
  - Navigation
  - Accessibility
  - Support features
  - Special functionality
  - Visual design

Analyzing the Data

- Quantitative - Such as completion rates, time on task
  - Analyses depend on sample sizes, data distribution (normal vs. non-normal)
  - Can be parametric or non-parametric
- Qualitative data - Descriptive, look for themes, triangulate with quantitative data for more comprehensive story
- How to interpret...how many errors = UT problem?

Additional Notes

- As with any data collection effort, always seek IRB approval for UTs. (Many times UTs are considered exempt/non-research due to their formative nature. Check with your IRB.)
- Try to add UT – even if you can only test 1-2 people!

For More Information

- Usability.gov
- Contact me: cpaulsen@concordevaluation.com
- Check out my upcoming AEA Coffee Break on December 15th