

think aloud practice

05.410, 05.610 & 05.819

Fall 2010



Human-Computer Interaction Institute

think-aloud evaluation is ...

asking participants to “think aloud” as they work...

... on a task you believe is interesting or important

... using a product/interface you are interested in improving

... while you observe (usually videotape)

... often in a “usability lab” equipped for this purpose



when to do a think-aloud

investigate your system ... is it ready for a think-aloud?

possible system forms:

- paper prototype
- simulation
- fragile prototype
- working system



define a framework for the test

what problem is this system trying to solve?

what level of support will our users have?

what types of use do we hope to evaluate?

what are the usability goals?



prioritize and select tasks

which tasks (intentions)?

- frequency
- critical
- representative

integration of small tasks?

duration of all tasks

will participants need training? it takes time



prepare

set up realistic situation ... do you need context?

write up task scenarios

rehearse, rehearse, rehearse

recruit appropriate participants ... do you need a screening system?



prepare participants

introduce yourself

ask them if they are willing to hear your “pitch” for participating in a study

describe the purpose in general terms

explain the terms of the study and get consent

give them consent form & get signature

ask them background questions



ethics & human participants

we test designs — not participants
never call this user testing

voluntary participation

- always make it OK for people to quit
- be sensitive to the many ways people express a desire to quit

maintain anonymity

- store the data under a code number or pseudonym
- store participant's name separately from their data



ethics & human participants

informed consent

laws and the irb (institutional review board)

CMU's compliance office:

<http://www.cmu.edu/osp/regulatory-compliance/human-subjects.html>



conduct the observation

introduce the observation phase

instruct them on how to think aloud: non-computer

instruct them on how to think aloud: computer

final instructions

- you won't be able to answer questions, but if questions cross their mind, say them aloud
- if you forget to think aloud, I'll say

"please keep talking"



analyze think-aloud data

insights come from words and actions ... don't just use a transcript

establish criteria for critical incidents

record critical incidents on Usability Aspect Report (UAR) forms



critical incident technique

By an **incident** is meant any *observable human activity* that is *sufficiently complete in itself* to permit inferences and predictions to be made about the person performing the act.

To be **critical**, an incident must occur in a situation where the *purpose or intent* of the act seems fairly clear to the observer and where its *consequences* are sufficiently definite to leave little doubt concerning its effects.

Such **incidents** are defined as *extreme behavior*, either outstandingly effective or ineffective with respect to attaining the general aims of the activity.

Flanagan, (1954), Psychological Bulletin, 51 (4), 327-358.



our criteria for bad incident

1. The user articulated a goal and does not succeed in attaining that goal within 3 minutes (then the experimenter steps in and shows him or her what to do--the next step).
2. The user articulates a goal, tries several things or the same thing over again (and then explicitly gives up).
3. The user articulates a goal and has to try three or more things to find the solution.
4. The user accomplishes the task, but in a suboptimal way
5. The user does not succeed in a task. That is, when there is a difference between the task the user was given and the solution the user produced.
6. The user expresses hesitation, surprise.
7. The user expresses some negative affect or says something is a problem.
8. The user makes a design suggestion (don't ask them to do this, but sometimes they do this spontaneously as they think-aloud).



our criteria for a good incident

- 9. The user expressed some positive affect or says something is really easy.
- 10. The user expresses happy surprise.
- 11. Some previous analysis has predicted a usability problem, but this user has no difficulty with that aspect of the system



what to look/listen for

analyst's mental framing: what is a problem for this system?

expression of intention or goal

actions that imply a goal or strategy

recognition of meeting or not meeting a goal

Q1: does it feel *sufficiently complete* (whole)?

Q2: does it seem extreme (effective or ineffective)



Usability Aspect Report (UAR)

Problem/Good Aspect:

"outstandingly effective or ineffective"

Name:

about the issue and not a proposed solution

Evidence:

"observable human activity"

Criterion:

from our list

Explanation:

"permit inferences and predictions"

...



example UAR 1

Name: System doesn't treat "1" and "one" interchangeably

Evidence:

What the user said: "My destination is One Market Street" [1:15]. "This is not right, I don't think this is what I want to do." [2:50] (on "Multiple Search Results" page).

What the user did: After receiving the "no exact match" error message, the user hits the Back button to return to the trip planner and enters "1" instead of "One."

What the system did: The system presented the user with a "Multiple Search Results" page, with the error message "We found some locations similar to your destination entry, but no exact match." After the user returns to the trip planner and enters "1," the system returns a bus route.

Criterion:

- 4. The user accomplishes a goal, but in a suboptimal way OR
- 7. The user expresses some negative affect or says something is a problem.

Explanation:

The user clearly has a goal of searching for One [1] Market Street, but the system returns an error message when the user attempts to achieve that goal by entering "One" instead of "1" in the address location. This meets criteria 4 (suboptimal) because the user has to enter the address twice, and meets criteria 7 (says something is a problem) because the user says "This is not right"



UAR name slot

Should name a PROBLEM, not a solution (e.g., font is too small to read, NOT change font to 14 point)

Overly narrow (e.g., “PA not listed” when there is nothing special about PA and other states besides PA are not listed)

Too broad, not distinctive (e.g., “User can’t find item”)



evidence

what's on the screen or what's presented aurally

what the user says

what the user does

what the system does in response to what the user does

framenumbers [pointer to evidence]



explanation

what you infer the user's goal was from what they said and/or did

what you infer the user was thinking from what the user said and/or did



let's practice



example UAR 2

Name: Select a day menu is inconsistent: mixes “today” and “tomorrow” with actual week names

Evidence:

What the user did: Opens the menu, checks the task to determine goal (“I want to do a Wednesday morning, according to the task”), [1:30] scrolls down over Today and Tomorrow to Thursday and then back up to Today, hesitates for a moment, [1:33], says “So.. today is Tuesday ... [hesitates] I think ... So tomorrow is Wednesday”, [1:38], selects Wednesday.

Criterion:

6. The user expresses confusion over how to achieve a goal

Explanation:

The user clearly states a goal of selecting Wednesday morning. His hesitation and the fact that he moves the mouse back and forth over the correct menu entry (“tomorrow”) indicate confusion (criteria 6)

