

Qualitative Evaluation Techniques

How to quickly evaluate prototypes by observing people's use of them

How specific methods can help you discover what a person is thinking about as they are using your system

Evan Golub / Ben Bederson / Saul Greenberg

Qualitative methods for usability evaluation

Qualitative:

- produces a description, usually in non-numeric terms
- may be subjective

Methods

- Introspection
- Direct observation
 - simple observation
 - think-aloud
 - constructive interaction
- Query via interviews and questionnaires

Evan Golub / Ben Bederson / Saul Greenberg

The Introspection Method (Walkthrough)

designer tries the system (or prototype) out

- does the system “feel right”?
- most common evaluation method



Problems

- not reliable as completely subjective
- not valid as introspector is a non-typical user

Also - Intuitions and introspection are often wrong!

Evan Golub / Ben Bederson / Saul Greenberg

Direct observation

Evaluator observes and records users interacting with design/system

- in lab:
 - user asked to complete a set of pre-determined tasks
 - a specially built and fully instrumented usability lab may be available
- in field:
 - user goes through normal duties

Excellent at identifying gross design/interface problems

Validity/reliability depends on how controlled/contrived the situation is...

Three general approaches:

- simple observation
- think-aloud
- constructive interaction

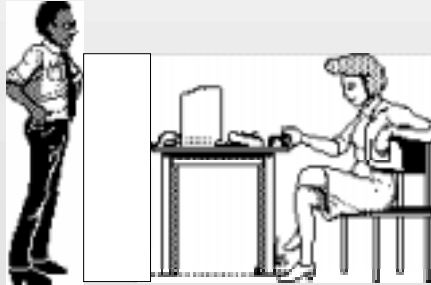
Evan Golub / Ben Bederson / Saul Greenberg

Simple Observation Method

User is given the task, and evaluator just watches the user

Problem

- does not give insight into the user's decision process or attitude



Evan Golub / Ben Bederson / Saul Greenberg

The Think Aloud Method

Subjects are asked to say what they are thinking/doing

- what they believe is happening
- what they are trying to do
- why they took an action
- Gives insight into what the user is thinking

Problems

- awkward/uncomfortable for subject (thinking aloud is not normal!)
- “thinking” about it may alter the way people perform their task
- hard to talk when they are concentrating on problem

Most widely used evaluation method in industry

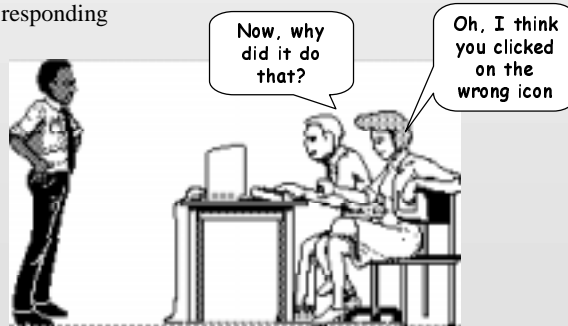


Evan Golub / Ben Bederson / Saul Greenberg

The Constructive Interaction Method

Two people work together on a task

- normal conversation between the two users is monitored
 - removes awkwardness of think-aloud
- Variant: Co-discovery learning
 - use semi-knowledgeable “coach” and naive subject together
 - make naive subject use the interface
- results in
 - naive subject asking questions
 - semi-knowledgeable coach responding
 - provides insights into thinking process of both beginner and intermediate users



Evan Golub / Ben Bederson / Saul Greenberg

Recording observations (make sure you ask permission)

How do we record user actions during observation for later analysis?

- if no record is kept, evaluator may forget, miss, or mis-interpret events
- paper and pencil
 - primitive but cheap
 - evaluators record events, interpretations, and extraneous observations
 - hard to get detail (writing is slow)
 - coding schemes help...
- audio recording
 - good for recording talk produced by thinking aloud/constructive interaction
 - hard to tie into user actions (ie what they are doing on the screen)
 - hard to search through later
- video recording
 - can see and hear what a user is doing
 - one camera for screen, another for subject (picture in picture)
 - can be intrusive during initial period of use
 - generates too much data



Evan Golub / Ben Bederson / Saul Greenberg

Coding scheme example...

tracking a person's activity in the office

s = start of activity
e = end of activity

Time	Desktop activities			Absences		Interruptions	
	working on computer	working on desk	initiates telephone	away from desk but in room	away from room	person enters room	answers telephone
9:00	s						
9:02	e					s	
9:05					s	e	
9:10			s		e		
9:13							



Evan Golub / Ben Bederson / Saul Greenberg

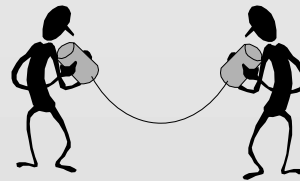
Querying Users via Interviews

Excellent for pursuing specific issues

- vary questions to suit the context
- probe more deeply on interesting issues as they arise
- good for exploratory studies via open-ended questioning
- often leads to specific constructive suggestions

Problems:

- accounts are subjective
- time consuming
- evaluator can easily bias the interview
- prone to rationalization of events/thoughts by user
 - user's reconstruction may be wrong



Evan Golub / Ben Bederson / Saul Greenberg

Structured Interviews

Plan a set of central questions

- could be based on results of user observations
- gets things started
- focuses the interview
- ensures a base of consistency

Try not to ask leading questions!

“Now that was easy, wasn’t it?”

“How hard would you say this task was?”

Start with individual discussions to discover different perspectives, and continue with group discussions

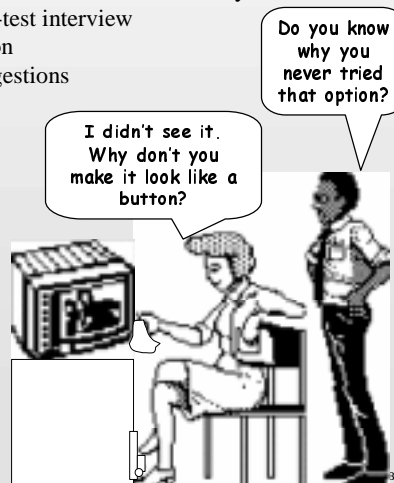
- the larger the group, the more the universality of comments can be ascertained
- also encourages discussion between users

Evan Golub / Ben Bederson / Saul Greenberg

Retrospective Testing

Post-observation interview to clarify events that occurred during system use

- perform an observational test
- create a video record of it
- have users view the video and comment on what they did
 - excellent for grounding a post-test interview
 - avoids erroneous reconstruction
 - users often offer concrete suggestions



Ben Bederson / Saul Greenberg

Evaluation through query continued: Methods

Questionnaires / Surveys

- preparation “expensive,” but administration cheap
 - can reach a wide subject group (e.g. mail)
- does not require presence of evaluator
- results can be quantified
- only as good as the questions asked
- often has low return rate - what’s in it for them?

- QUIS - Questionnaire for User Interface Satisfaction
 - Shneiderman text has a version of it



Evan Golub / Ben Bederson / Saul Greenberg

Querying Users via Questionnaires / Surveys

How

- establish the purpose of the questionnaire
 - what information is sought?
 - how would you analyze the results?
 - what would you do with your analysis?
- typically will not ask questions whose answers you will not use
 - this is unlike many other types of surveys you may have discussed in your psychology class
- determine the audience you want to reach
 - typical survey: random sample of between 50 and 1000 users of the product
- determine how would you will deliver and collect the questionnaire
 - on-line for computer users
 - web site with forms
 - surface mail (including a pre-addressed reply envelope gives far better response)
- determine the demographics
 - e.g. computer experience

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (I)

Open-ended questions

- asks for unprompted opinions
- good for general subjective information
 - but difficult to analyze rigorously

eg: **Can you suggest any improvements to the interfaces?**

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (II)

Closed questions

- restricts the respondent's responses by supplying alternative answers
- makes questionnaires a chore for respondent to fill in
- can be easily analyzed
- but watch out for hard to interpret responses!
 - alternative answers should be very specific

Do you use computers at work:

often sometimes rarely

-vs-

In your typical work day, do you use computers:

- over 4 hrs a day
- between 2 and 4 hrs daily
- between 1 and 2 hrs daily
- less than 1 hr a day

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (III)

Scalar

- ask user to judge a specific statement on a numeric scale
- scale usually corresponds with agreement or disagreement with a statement

Characters on the computer screen are:

hard to read easy to read

1 ② 3 4 5

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (IV)

Multi-choice

- respondent offered a choice of explicit responses

How do you most often get help with the system? (tick one)

- on-line manual
- paper manual
- ask a colleague

Which types of software have you used? (tick all that apply)

- word processor
- data base
- spreadsheet
- compiler

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (V)

Ranked

- respondent places an ordering on items in a list
- useful to indicate a user's preferences
- forced choice

Rank the usefulness of these methods of issuing a command

(1 most useful, 2 next most useful..., 0 if not used)

 2 command line

 1 menu selection

 3 control key accelerator

Evan Golub / Ben Bederson / Saul Greenberg

Styles of Questions (VI)

Combining open-ended and closed questions

- gets specific response, but allows room for user's opinion

It is easy to recover from mistakes:

disagree agree comment: *the undo facility is really helpful*
1 2 3 (4) 5

Evan Golub / Ben Bederson / Saul Greenberg

What you now know

Observing a range of users use your system for specific tasks reveals successes and problems

Qualitative observational tests are quick and easy to do

Several methods reveal what is in a person's head as they are doing the test

Particular methods include

- Conceptual model extraction
- Direct observation
 - simple observation
 - think-aloud
 - constructive interaction
- Query via interviews, retrospective testing and questionnaires
- Continuous evaluation via user feedback and field studies

Evan Golub / Ben Bederson / Saul Greenberg