

Qualitative Evaluation Techniques

Why evaluation is crucial to interface design

General approaches and tradeoffs with the different approaches to evaluation

The role of ethics

Learning how to quickly debug and evaluate prototypes by observing people using them

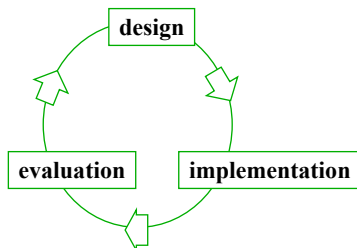
Specific evaluation methods helps you discover people's thoughts and motivations as they are using your system

James Tam

Why Bother With Evaluation?

Tied to all parts of the usability engineering lifecycle

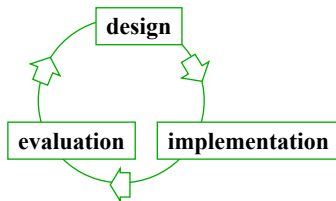
- Pre-design
 - Investing in new expensive system requires proof of viability
- Initial design stages
 - Develop and evaluate initial design ideas with the user



James Tam

Why Bother With Evaluation (2)?

- Iterative design
 - Does the system match the user's task requirements?
 - Are there any specific problems with the design?
 - Can users provide feedback to modify the design
- Acceptance testing
 - Verify that the system meets expected performance criteria: ease of learning, usability, user's attitude
 - E.g., A first time user will take between one and three minutes to learn how to withdraw \$50 from the ATM.



James Tam

Approaches: Naturalistic

Observation occurs in realistic setting

- Real life
- Problems
 - It may be difficult to arrange and to conduct
 - It may not always possible to replicate results

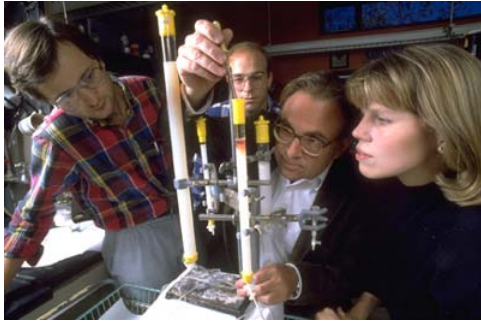


James Tam

Approaches: Experimental

Experimental

- Classical lab study
- Study relations by manipulating one or more *independent* variables
 - Experimenter controls all environmental factors (nothing else is different)
- Observe effect on one or more *dependent* variables



James Tam

Tradeoffs: Natural Vs. Experimental

Internal validity

- Do you measure what you set out to measure (correctness)

External validity

- The degree to which results can be generalized to other situations (realism)

	Naturalistic	Experimental
Internal validity	Low	High
External validity	High	Low

James Tam

(External) Validity Concerns

Does the test measure something of relevance to usability of real products in real use outside of lab?

- Some typical reliability problems of testing vs. real use
 - Non-typical users tested
 - Tasks are not typical tasks
 - Physical environment different
 - quiet lab vs.. very noisy open offices vs. interruptions
 - Social influences different
 - motivation towards experimenter vs. motivation towards boss



Partial Solution

- Use real users
- Use tasks derived from a task-centered approach to system design
- Test in an environment similar to real situation



James Tam

How Many Participants To Test

Would the same results be achieved if the test were repeated?

- Reliability of data

Problem: individual differences:

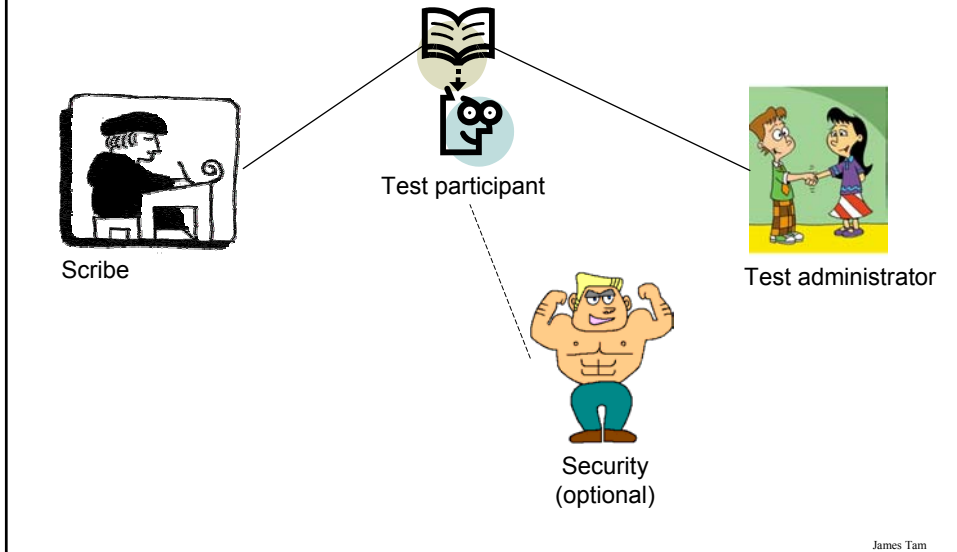
- The best user 10x faster than slowest
- The best 25% of users ~2x faster than slowest 25%



Partial Solution

- Get a reasonable number and range of test participants

Test Roles



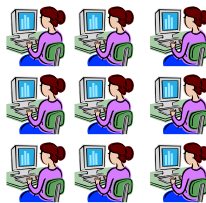
Test Procedure

I) Run a pilot study



- "A practice run" of the test
- Purpose: To debug the test
- Results: Used to improve the test

II) Run the main test



- Running the test "for real"
- Purpose: To debug the interface
- Results: Used to improve the interface

Test Procedure (2)

1. Preparation
2. Introduction
3. Running the system
4. Debriefing

James Tam

Ethics



James Tam

Ethics

Testing can be a distressing experience

- People feel pressure to perform so errors are inevitable
- This can result in:
 - Feelings of inadequacy
 - Competition with other test participants



Golden rule

- Test participants should always be treated with respect

James Tam

Managing Participants In An Ethical Manner

Before the test

- Don't waste the person's time
 - Use pilot tests to debug experiments, questionnaires etc
 - Have everything ready before the participant shows up
 - Try it out yourself one more time
- Make participants feel comfortable
 - Emphasize that it is the system that is being tested, not the person
 - Acknowledge that the software may have problems
 - Let participants know they can stop at any time
- Maintain privacy
 - Tell the participant that individual test results will be kept completely confidential
- Inform the participant
 - Explain any monitoring that is being used
 - Answer all of the person's questions (but avoid biasing them)
- Only use volunteers
 - Typically the test participant must sign an informed consent form



James Tam

Managing Participants In An Ethical Manner

During the test

- Don't waste the person's time
 - Never have the user perform unnecessary tasks
- Make test participants comfortable
 - Try to give the person an early success experience
 - Keep a relaxed atmosphere in the room
 - Have coffee, breaks, etc
 - Hand out test tasks one at a time
 - Never indicate displeasure with the person's performance
 - Avoid disruptions
 - Stop the test if it becomes too unpleasant
- Maintain privacy
 - This class: Only show test results to people when it is essential (TA and course instructor)
 - Actual practice: Do not allow the participant's management to observe the test



James Tam

Managing Participants In An Ethical Manner

After the test

- Make the person feel comfortable
 - e.g., state that the participant has helped you find areas of improvement
- Inform the participant
 - Answer particular questions about the experiment that could have biased the results before
- Maintain privacy
 - Never report results in a way that individuals can be identified
 - Only show videotapes outside the research group with the participant's permission



James Tam

Canon
 Fax-B320
 Bubble Jet Facsimile

ON LINE
 0
 PRINTER ERROR
 0
 PRINTER

SHQ
 HS HQ
 0 0
 PRINT MODE

PRINTER
 INTERFACE



1 2 3 CODED DIAL /DIRECTORY

4 5 6 R

7 8 9 Pause

* 0 #

< >
 v ^

memory delayed delayed
 trans trans polling polling

01 02 03 04

confd relay
 trans broadca report

05 06 07 08

+ D.T. Tone

09 10 11 12

space clear

13 14 15 16

HOLD

James Tam

Discount Usability Evaluation

Low cost methods to gather usability problems

- Approximate: Capture most large and many minor problems

How?

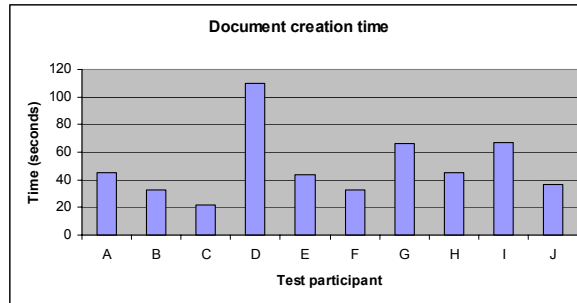
- Quantitative
- Qualitative

James Tam

Quantitative Approach For Usability Evaluation

Description of approach:

- Measure something of interest in user actions
- Count, log, speed, error rate



James Tam

Qualitative Methods For Usability Evaluation

Description of approach:

- Data gathering
 - Observe the actions of the user
 - Gather opinions from the user
- Produces a description, usually in non-numeric terms
- May be quite subjective

Methods

- Inspection
- Extracting the conceptual model
- Direct observation
 - Simple observation
 - Think-aloud
 - Constructive interaction
- Query via interviews and questionnaires
- Continuous evaluation via user feedback and field studies



Star Trek IV: The Voyage Home © Paramount Pictures

James Tam

The Inspection Method

Designer tries the system (or prototype) out

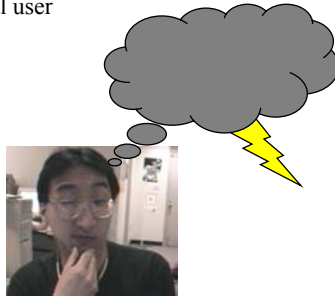
- Does the system “feel right”?

Benefits

- Can probably notice some major problems in early versions during every day use

Problems

- Low reliability rate as completely subjective
- Low level of validity as inspector is a non-typical user
- Intuitions and introspections are often wrong



James Tam

Extracting The Conceptual Model

Show the user static images of:

- The paper prototype *or*
- Screen snapshots *or*
- Actual system screens during use

Have the user try to explain

- What all elements are
- What they would do to perform a particular task
- How they think that the system works

Initial vs.. formative conceptual models

- **Initial:** How person perceives a screen the very first time it is viewed
- **Formative:** The same, except after the system has been used for a while

James Tam

Extracting The Conceptual Model (2)

This approach is:

- Good for eliciting people's understanding before & after use
- Requires active intervention by evaluator, which can get in the way

James Tam

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

This approach is:

- Validity/reliability depends on how controlled/contrived the situation is
- Excellent at identifying gross design/interface problems

Three general approaches:

- Simple observation/Silent observer
- Think-aloud
- Constructive interaction

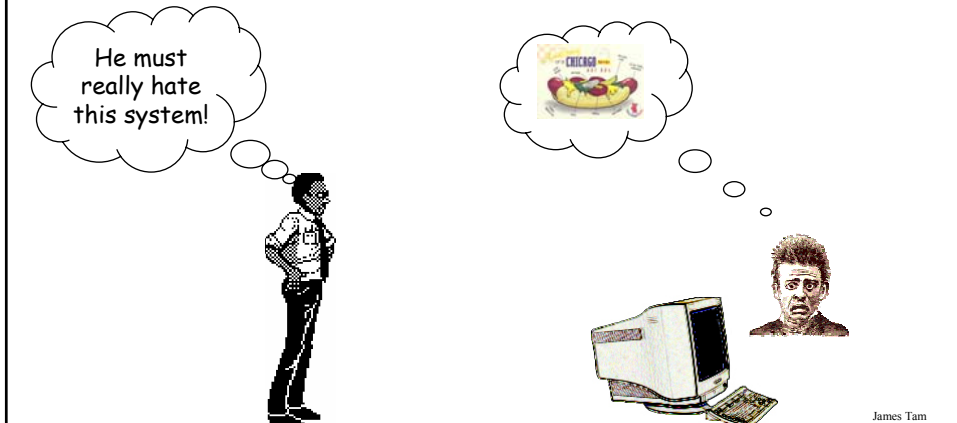
James Tam

Simple Observation Method

Person is given the task, and evaluator silently just watches while employing “The Silent Observer” technique.

Problem

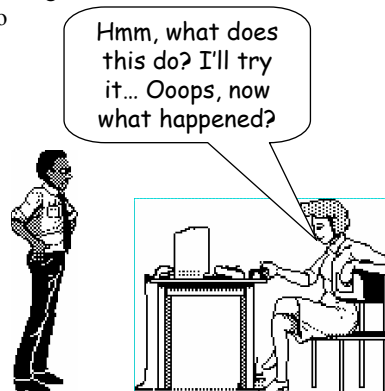
- Does not give insight into the person’s decision process or attitude



The Think Aloud Method

Test participants are asked to say what they are **thinking/doing**

- Gives insight into what the person is thinking
 - What they believe is happening
 - What they are trying to do
 - Why they took an action



The Think Aloud Method (2)

Problems

- Awkward/uncomfortable for person (thinking aloud is not normal!)
- Hard to talk when they are concentrating on problem
- “Thinking” about it may alter the way people perform their task (could improve *or* degrade performance)
- Certain situations may prohibit the use of this technique

Most widely used “formal” evaluation method in industry

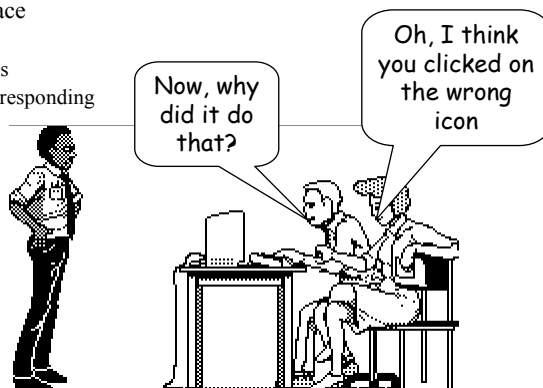
James Tam

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 novice user together
 - Only novice uses the interface
 - Results in:

Novice user asking questions
Semi-knowledgeable coach responding
Provides insights into
thinking process of both
user groups



James Tam

Querying People Via Interviews

Use a set of pre-created questions

- Gets things started
- Focuses the interview
- Ensures a base of consistency
- Be sure to follow interesting leads rather than bulldozing through question list

Adding additional questions

- Could be based on results of user observations

Don't forget

- Balance each question
- Avoid bias
 - Try not to ask leading questions



THE INTERVIEWER DIDN'T
EAT LUNCH.

James Tam

Issues Associated With Interviews

Excellent for pursuing specific issues

- Flexible
 - You can vary questions to suit the context
- Provides a rich depth of data
 - Probe more deeply on interesting issues as they arise
 - Often leads to specific constructive suggestions

Problems:

- Time consuming
- Evaluator can easily bias the interview
- Requires a skilled and/or experienced interviewer
- Accounts are subjective
- Prone to rationalization of events/thoughts by person
 - Reconstruction may be wrong

James Tam

Group Discussions

- **Start with individual discussions to discover different perspectives, and then continue with group discussions**
- **Increasing group size may increase the universality of the comments**
- **May encourage cross discussions**



James Tam

Retrospective Testing

A special type of interviewing technique that was developed in order to address the weaknesses of traditional interviews.

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

Approach:

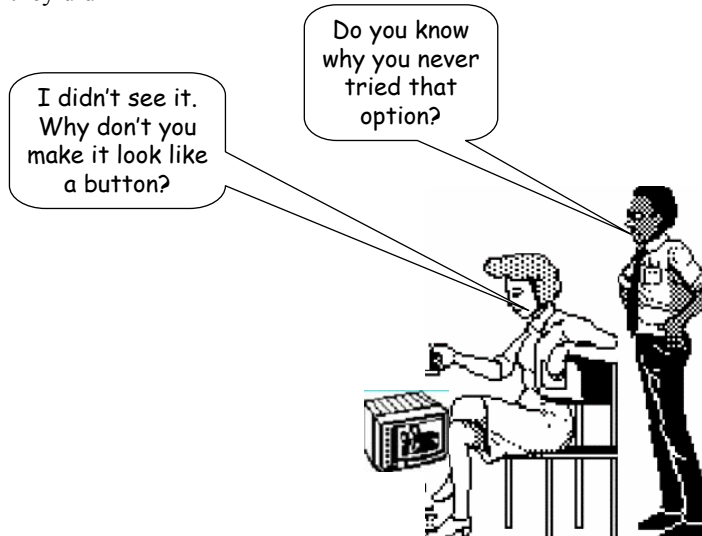
1. Perform an observational test while recording the session on video



James Tam

Retrospective Testing (2)

2. Watch the video with the users and encourage them to comment on what they did



Retrospective Testing (3)

Benefits

- Excellent for grounding a post-test interview
- Avoids erroneous reconstruction
- Users often offer more concrete suggestions
- Unlike the silent observer approach it provides insights into what the person is thinking/feeling and it can be used when thinking aloud is not possible

Drawbacks

- Much like traditional interviews it can be very time consuming

Querying People Via Questionnaires And Surveys

Questionnaires / Surveys

- Written queries for usability information

Benefits

- Administration cheap
 - Can reach a wide test group (e.g., mail)
- Administration requires little training
- Anonymous



Drawbacks

- Preparation “expensive” – although this may be balanced off by the administrative savings
- Inflexible

James Tam

Querying People Via Questionnaires / Surveys (2)

Approach for all types

- Establish the purpose of the questionnaire
 - What information is sought?
 - How would you analyze the results?
 - What would you do with your analysis?

http://www.cpsc.ucalgary.ca/~tamj/481/assignments/usability/questionnaire_tips.html

- Do not ask questions whose answers you will not use!
 - e.g. How old are you?
- Determine the audience you want to reach
 - Typical survey: random sample of between 50 and 1000 users of the system
- Determine how you will deliver and collect the questionnaire
 - On-line for computer users (e.g., web site with fill-in forms)
 - Surface mail
 - Be sure to include a pre-addressed reply envelope to get a far better response rate

James Tam

Querying Users Via Questionnaires / Surveys (3)

- Determine the demographics
 - e.g., computer experience

James Tam

Style Of Questions

Open-ended questions

- Asks for unprompted opinions
- Good for general subjective information but difficult to analyze rigorously

e.g., Can you suggest any improvements to the interface?

James Tam

Style Of Questions

Closed-ended questions

- Restricts the respondent's responses by supplying alternative answers
- Data is more narrow (less rich but can be easily analyzed)
- But watch out for hard to interpret responses - alternative answers should be very specific
- Types: scalar, multiple choice, ranked
- Examples:

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

James Tam

Closed-Ended Questions: Scalar

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

James Tam

Closed-Ended Questions: Multiple Choice

Multi-choice

- Respondent offered a choice of explicit responses

How do you most often get help with the system? (Check only one category)

- On-line manual
- Paper manual
- Ask a colleague

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

- Word processor
- Data base
- Spreadsheet
- Compiler

James Tam

Closed-Ended Questions: Ranked

Ranked

- Respondent places an ordering on items in a list
- Useful to indicate a user's preferences
- Forces a choice

Rank the usefulness of these methods of giving a command to the computer

(1 = Most useful, 2 = Next most useful..., 0 = Not used)

- 2 Command line
- 1 Menu selection
- 3 Control key accelerator

James Tam

Mixing Questionnaire Styles

Combining open-ended and closed-ended questions

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

It is easy to recover from mistakes:

Disagree

1

2

3

4

5

Agree

Comment: *The undo facility is really helpful*

James Tam

Interviews Vs. Questionnaires: Summary Of The Pros And Cons

- Preparation time
- Unanticipated/unexpected events
- Depth of information
- Analysis time

James Tam

Recording Observations

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 or having a second observer may be helpful:

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							

↓

James Tam

Recording Observations (2)

- Audio recording
 - Good for recording the dialog produced by thinking aloud/constructive interaction
 - Hard to tie into user actions (i.e., what they are doing on the screen)
- Video recording
 - Can see and hear what a user is doing
 - One camera for screen, another for test user (picture in picture)
 - Can be intrusive during initial period of use

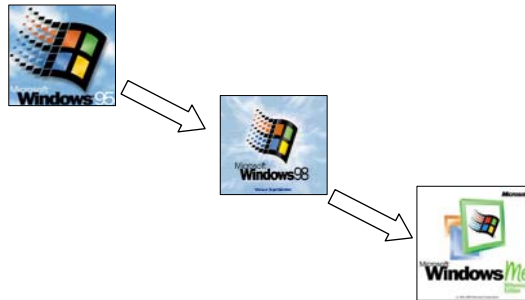


James Tam

Continuous Evaluation

1) Developers monitor system while it's actually being used

- Usually done in later stages of development
 - i.e., Beta releases, delivered system
- Good for finding real-world problems
- Problems can be fixed in the next release



Windows is the property of Microsoft Corporation

James Tam

Continuous Evaluation (2)

2) Users can provide feedback

- Email
- Special built-in gripe facility (web site, bulletin board)
- Telephone hot line
- Help desks
- Suggestion boxes



Best combined with trouble-shooting facility

- Users always get a response (solution?) to their problem



James Tam

Continuous Evaluation (3)

3) Case/field studies

- Careful study of “system usage” at the site
- Good for seeing “real life” use
- Can be informal or more rigorous qualitative approaches can be attempted



James Tam

What You Now Know

Evaluation is crucial for designing, debugging, and verifying interfaces

There is a tradeoff in naturalistic vs. experimental approaches

- Internal and External validity

The number and range of test participants employed will effect the reliability of your results

Test participants *must* be treated with respect

- The study should be guided by ethical rules of behavior

James Tam

What You Now Know (2)?

Observing a range of users use your system for specific tasks reveals many 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

- Inspections
- Conceptual model extraction
- Direct observation
 - Simple observation
 - Think-aloud
 - Constructive interaction (Co-discovery learning)
- Query via interviews, retrospective testing and questionnaires
- Continuous evaluation via user feedback and field studies

James Tam

