



UNIVERSITY OF SCIENCE  
HO CHI MINH CITY

CTT534 – Thiết Kế Giao Diện  
HK II 2013 – 2014

# UI Design Evaluation

Some slides adapted from MIT CS Course  
6.813/6.831 and USC's CSCI 588

# Outline

- Overview of user interface design evaluation
- UI evaluation methods
  - Expert
    - Heuristic evaluation
  - User
    - Field studies
    - Controlled experiment
    - Formative evaluation

# Why, what, where, when to evaluate

- Iterative design and evaluation is a continuous process that examines:
  - Why: to check that users can use the product and that they like it
  - What: a conceptual design, early prototypes and later, more complete prototypes
  - Where: in natural and laboratory settings
  - When: throughout design process; finished products can be evaluated to collect information to inform new products

# Throughout the Usability Engineering Lifecycle

- Conceptual design
  - developing and evaluating initial concept design ideas with the users
- Early prototype design
  - developing continuous process with the users
    - does system behavior match the user's task requirements?
    - are there specific problems with the design?
- Complete prototype design
  - acceptance testing to verify the system meets expected user performance criteria

# Purpose of evaluation

- Assess the effect of the interaction on the user
  - usability and user satisfaction
- Assess the extent of system's functionality
- Identify specific problems in the design
  - aspects of design that cause unexpected results or confusion

# Methods of evaluation

- Heuristic evaluation
  - Done by experts
- User tests: done by users
  - Field studies
  - Controlled experiment
  - Formative evaluation

# Outline

- Overview of user interface design evaluation
- UI evaluation methods
  - Expert
    - **Heuristic evaluation**
  - User
    - Field studies
    - Controlled experiment
    - Formative evaluation

# Usability heuristics/guidelines

- Many heuristics to base on
  - Nielsen's 10 heuristics
  - Norman's rules from Design of Everyday Things
  - Shneiderman's Eight Golden Rules
- Heuristics help
  - Designers to choose design alternatives
  - Help evaluators find problems and assess interfaces (hence Heuristics Evaluation)



# UI design principles we've studied

- Learnability/Memorability – L for short
- Visibility – V
- Simplicity – S
- User control – UC
- Error handling – ER
- Efficiency – E
- Graphic design – GD

# Nielsen's heuristics

- Match the real world (L)
- Consistency and standards (L)
- Help and documentation (L)
- User control and freedom (UC)
- Visibility of system status (V)
- Flexible and efficiency (EF)
- Error prevention (ER)
- Recognition, not recall (ER)
- Error reporting, diagnosis, and recovery (ER)
- Aesthetic and minimalist design (GD)

# Norman's rules

- Affordances (L)
- Natural mapping (L)
- Visibility (V)
- Feedback (S)

# Shneiderman's eight golden rules

1. Consistency (L)
2. Shortcuts (EF)
3. Feedback (V)
4. Dialog closure (V)
5. Simple error handling (ER)
6. Reversible actions (UC)
7. Put users in control (UC)
8. Reduce short-term memory load (ER)

# Heuristic evaluation

- Invented and evaluated by Jakob Nielsen
  - ❑ Cheaper and more effective than alternative methods
- An inspection technique
- Often performed by experts
- Basic steps
  - ❑ Review/inspect UI
  - ❑ Compare it against certain heuristics
  - ❑ Document usability problems
  - ❑ Explain and justify problems against heuristics

# Guidelines for heuristic evaluation-1

- Justify every problem with heuristics (one or many)
  - “Dialog A has button ‘Hủy’ while the same button on dialog B is labeled ‘Đóng’” (Consistency)
  - Don’t just say “I don’t like the layout”
- List every problem
- Go through the interface multiple times
- Don’t have to limit to the 10 Nielsen’s heuristics
  - Also, not all heuristics are applicable for a user interface

# Guidelines for heuristic evaluation-2

- Use multiple evaluators
  - ❑ More people find more and different problems than one person
  - ❑ Problems can be duplicated
  - ❑ Nielsen recommends 3-5 evaluators
- Use heuristic evaluation in addition to user testing
  - ❑ Each method finds different problems
  - ❑ Heuristic evaluation is cheaper

# Let's evaluate the U. of S's website



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
TP. HO CHI MINH

ĐẠI HỌC QUỐC GIA TP.HCM  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

Tiếng Việt English



Thứ Tư, 18/04/2012, 07:49 (GMT+7)

Tìm kiếm

- Trang Chủ
- Giới Thiệu
- Đào tạo
- Tuyển sinh
- Khoa học CN-QHQT
- Công tác Sinh viên
- Tổ chức hành chính
- Đoàn thể
- Thư viện
- Các Khoa
- Các Phòng Thí Nghiệm
- Các Trung Tâm
- Các phòng ban

Chào mừng các bạn đến với website  
**Trường ĐH Khoa học Tự nhiên**



**Giới thiệu**

Năm 1996, Trường ĐH KHTN được chính thức thành lập theo quyết định 1236/GDBT của Bộ GD&ĐT ngày 30/3/1996 trên cơ sở tách ra từ Trường Đại học Tổng hợp TP.HCM để tham gia vào Đại học Quốc gia TP.HCM. Trường hiện có 09 Khoa; 15 trung tâm đào tạo, nghiên cứu khoa học- chuyển giao công nghệ. Hiện nay, Trường đang có 32...

Chi tiết...



**THÔNG TIN TUYỂN SINH 2012**

Tin Khoa học tự nhiên

**Thứ vị với sáng kiến "Bảng vẽ thông minh" của sinh viên ĐH KHTN**

16/04/2012, 10:04 (GMT+7)

Xuất phát từ niềm đam mê với các công nghệ về tương tác giữa người và máy tính, đồng thời nhận thấy việc đưa bề mặt cảm ứng vào môi trường giáo dục là hướng tiếp cận mới và cần thiết nhằm đổi mới các thiết bị dạy học phù hợp sao cho chuyển tải kiến thức hiệu quả nhất; các bạn sinh viên trường **ĐH Khoa học Tự nhiên TP.HCM** đã có sáng kiến làm ra bề mặt cảm ứng sử dụng tín hiệu hồng ngoại dựa vào thị giác máy tính và hệ thống phần mềm được viết trên framework WTF thay thế cho các dụng cụ dạy học truyền thống: phấn trắng, bảng đen.

...

Chi tiết...

**Các hoạt động khác**

- Sản xuất túi ni lông tự hủy từ bột mì (12/04/2012)
- Điều kiện tuyển thẳng vào trường ĐH KHTN kỳ tuyển sinh năm 2012 (06/04/2012)
- Xe năng lượng mặt trời chỉ 15 triệu đồng (04/04/2012)
- "Đêm hội Văn hóa" trường ĐH KHTN: Nơi tuyên dương những tấm gương SV tiêu biểu (29/03/2012)
- Đại hội lần V Hội Cựu chiến binh trường ĐH KHTN kết thúc thành công tốt đẹp (27/03/2012)
- ĐH KHTN: 113 Cử nhân ngành CNTT hệ vừa làm vừa học ra trường (26/03/2012)
- ĐH KHTN TP.HCM: Tuyển sinh khối A1 ba ngành bậc đại học năm 2012 (19/03/2012)
- ĐH KHTN: Nhiều hoạt động chào mừng Ngày Quốc tế Phụ nữ (09/03/2012)
- Phó thủ tướng Nguyễn Thiện Nhân thăm và làm việc với ĐH KHTN (08/03/2012)

**Lịch công tác**

**Hộp thư Cán bộ**

Hướng dẫn sử dụng  
Hộp thư Cán bộ

**Hộp thư Sinh viên**

**Liên hệ**

**Download Logo**

**Hình ảnh**

**Tuyển dụng**



**Hệ đào tạo từ xa qua mạng**

**Đăng nhập**

Tên tài khoản

Mật khẩu

Đăng nhập

☐ Lưu thông tin | Quên mật khẩu

Kích hoạt tài khoản | Hướng dẫn

**Thông báo đào tạo**

- Thông báo nộp hồ sơ học bổng Viettel năm 2011 (lần 2) (18/04/2012) new!
- Thông báo nghỉ học ngày 22/4/2012 tại cơ sở Linh Trung (do cúp điện) (18/04/2012) new!
- Thông báo nộp hồ sơ xét Tốt nghiệp Đại học, hệ Vừa học vừa làm, đợt tháng 5/2012 (18/04/2012) new!
- Thông báo v/v nhận bổ sung chứng chỉ ngoại ngữ để xét miễn thi ngoại ngữ trong kỳ thi tuyển sinh Sau đại học năm 2012 (17/04/2012) new!
- Thông báo v/v nhận bổ sung chứng chỉ ngoại ngữ để xét miễn thi ngoại ngữ trong kỳ thi tuyển sinh Sau đại học năm 2012 (17/04/2012) new!
- Thông báo nộp hồ sơ chương trình

TP. HO CHI MINH



# Evaluating prototypes

- Heuristic evaluation can also be used to evaluate
  - Sketches
  - Paper/computer prototypes
  - Demo code
- Certain heuristics are not applicable
  - Feedback
  - “Missing-element” problems
  - Help and documentation

# Formal evaluation process

## ■ Training

- Meeting to introduce the application and its user interface
- Explain user population, domain, scenarios

## ■ Evaluation

- Evaluators work independent to record problems and generate report

## ■ Severity rating

- Evaluators provide severity levels for problems found

## ■ Briefing

- Evaluators and design team discuss results and suggest solutions

# Severity level

- Different scales can be provided, such as
  - ❑ Cosmetic or trivial: not important, low priority
  - ❑ Minor: not serious, low priority
  - ❑ Major: quite serious, high priority
  - ❑ Critical: serious, high priority
- Factors to consider for rating
  - ❑ Frequency: how common the problem is found?
  - ❑ Impact: hard to resolve?
  - ❑ Persistence: repeated?

# Writing heuristic evaluations - 1

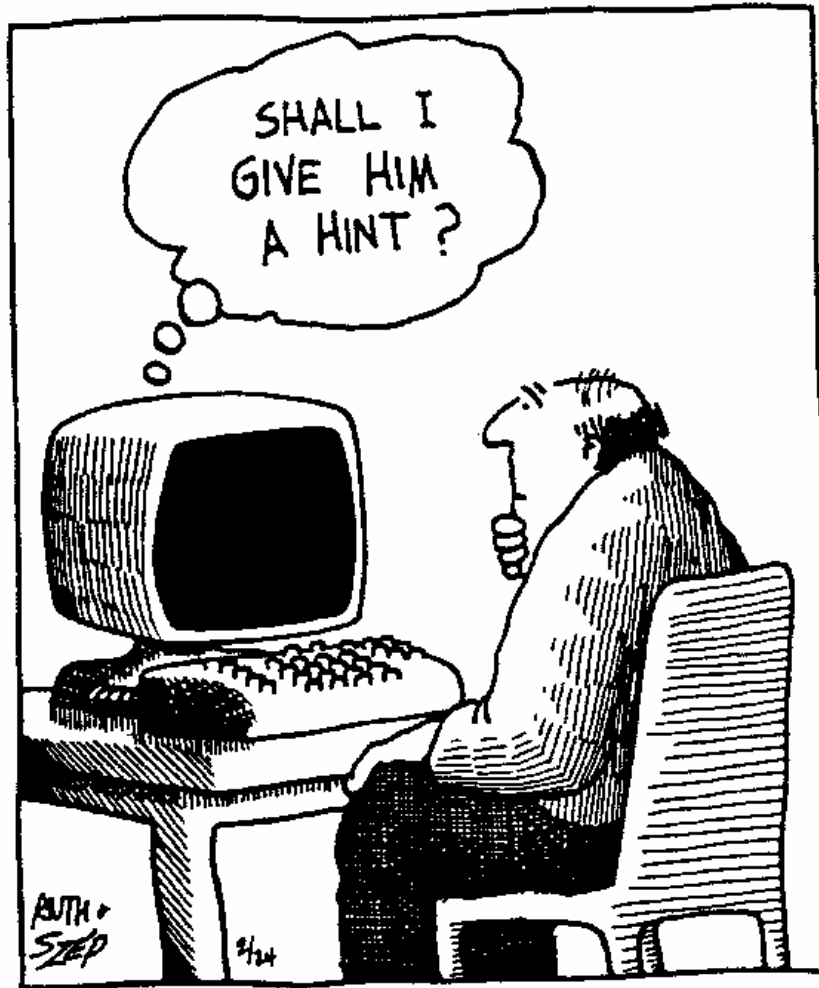
- Evaluators must provide clear descriptions of problems
  - To reduce time for both designers and evaluators
- Include both positive comments and criticisms
  - Emphasize more on criticisms and recommendations
- Don't be unnecessarily tough
- Be specific
  - Not specific: “text is unreadable”
  - Specific: “text has poor contrast”

# Writing heuristic evaluations - 2

- Evaluation report can include
  - Problem title
  - Problem description
  - Heuristic
  - Severity
  - Recommendations and screenshots (if any)

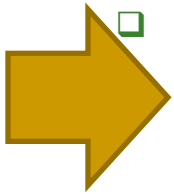
# Cognitive walkthrough

- Is an inspection technique that focuses on learnability
- Inputs
  - Prototype
  - Task
  - Sequence of actions
  - User analysis
- Evaluator asks questions while following sequence of actions
  - Will user find the action on the interface?
  - Will user know what to do?
  - Etc.



# Outline

- Overview of user interface design evaluation
- UI evaluation methods
  - Expert
    - Heuristic evaluation
  - User
    - **Field studies**
    - Formative evaluation
    - Controlled experiment





# Field studies

- Real use environments
  - Observe effect occurs in realistic setting
  - Can't tell how good UI is until the users use the products and that they satisfy
  - Hard to predict what real users will do
    - problems identified are those that are really raised when the user uses the software
  - Usability could be measured during the test by calculating the user performance

# Field studies (cont'd)

- Problems
  - ❑ cost/time consuming
  - ❑ may not generalize

# Procedure

1. Prepare test proposal
2. Choose participants
3. Select properties to be tested
4. Perform the tests
5. Measure the test results

# 1. Prepare test proposal

- Test proposal
  - test objective
  - test participants
  - test environment and materials
  - test methodology to be used
    - detailed descriptions of systems/tasks/properties being tested
  - test measures
    - test data to be collected and analyzed

## 2. Choose participants

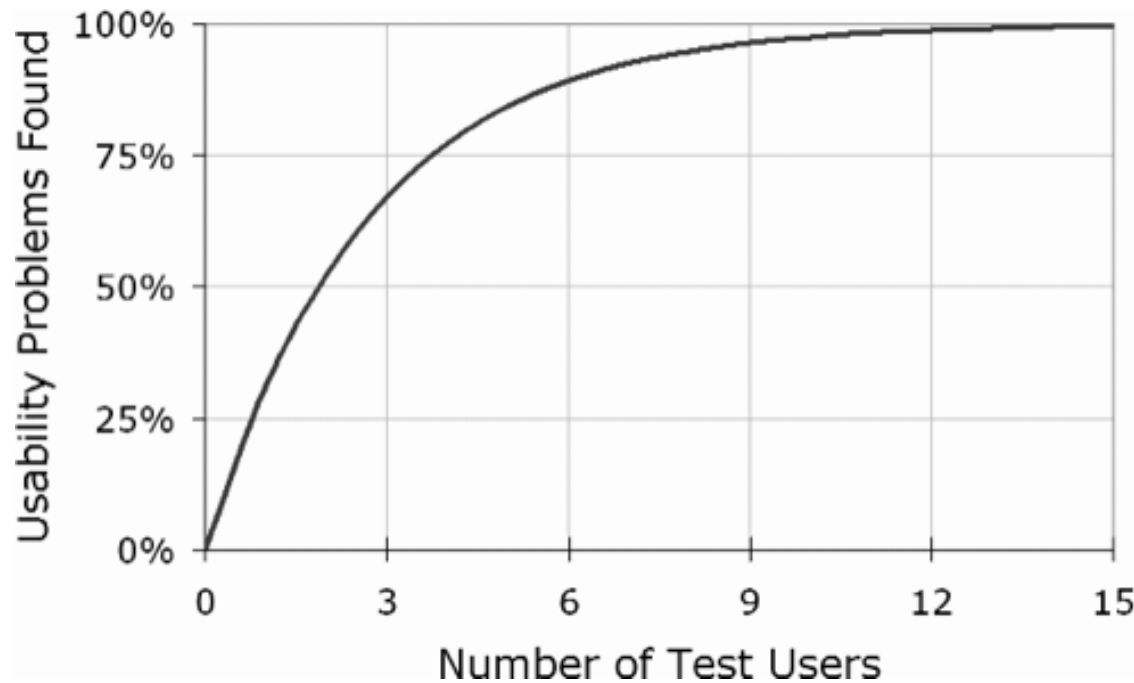
- Representative of target users
  - The users must be those targeted by the evaluated product
  - Relevant tasks, job-specific knowledge
- Approximate if needed
  - system intended for doctors?
    - get medical students
  - system intended for engineers?
    - get engineering students

## 2. Choose participants (cont'd)

- How many users are enough?
  - Too many users?
    - Cost issue - observing many users is time/cost consuming
  - Too few users
    - High variation
    - Subjective issue - individual differences matter
- Partial solution
  - reasonable number of users
    - big problems usually detected with handful of users
    - small problems and fine measures need many users

## 2. Choose participants (cont'd)

- How many users are enough?
  - J. Nielsen has shown that tests performed with 5 users raise at least 80% of the usability problems
  - 3 tests with 5 users is better than 1 test with 15 users



# Ethical considerations

- Users are human beings!
- Testing can be a distressing experience
  - pressure to perform, errors inevitable
  - feelings of inadequacy
- You have a responsibility to alleviate:
  - make voluntary testers with informed consent
  - make them feel comfortable, avoid pressure
  - let them know they can stop at any time
  - stress that you are testing the system, not them
  - make collected data as anonymous as possible
- Law often requires to get human subjects approval



# Ethical considerations (cont'd)

- The objective of test is to identify usability problem, not to measure user's capability

### 3. Select properties to be tested

- Define a precise objective for each test
  - Well-defined test instructions and objectives → test results are easy to interpret
- Test objective should reflect real tasks
- Avoid bending tasks in direction of what your design best supports
- Don't choose tasks that are too fragmented
- The objective is to evaluate the software, not the user!

# 4. Perform the tests

- Evaluation methods
  - Inspection
  - Direct observation
  - Query techniques
    - Interview/questionnaires/surveys

# Inspection

- Try system/prototype
  - does the system “feel right”?
- Benefits
  - can catch some major problems in early versions
  - often help the task centered walkthroughs
- Problems
  - not reliable as completely subjective

# Direct observation

- Evaluator observes/records users interacting with systems
- In usability laboratory
  - user completes set of pre-determined tasks
- In fields
  - user goes through normal duties
- Benefits
  - excellent at identifying gross design/interface problems
  - validity depends on how the situation is controlled



# Query techniques

- Evaluator asks user about opinion of the system
- The most common methods include
  - Interview
  - Questionnaire / surveys

© 1998 Randy Glasbergen  
www.glasbergen.com



"We also have some part-time positions available for people who only want to work 60 or 80 hours a week."

# Query techniques: Interview

- Evaluator questions user (on one-to-one basis)
  - based on prepared questions
    - Structured, unstructured, or both
  - informal, subjective and relatively cheap
- Good for pursuing specific issues with user
- Flexible as evaluator can probe more deeply into interesting issues that arise

# Interview (cont'd)

- Prepared a set of central questions
  - a few good questions gets things started
- Could be based on results of user observations
  - post-observation interview
- Let user responses lead follow-up questions
  - avoid user leading the test



# Interview (cont'd)

## ■ Post-observation Interview

- have the users perform an observational test
- create a video record of the test
- interview the users with the video record
  - users comment on the events that occurred during system use
    - – e.g. I used this button is because...
  - users often offer concrete suggestions
    - – e.g. you should make this button smaller...

# Interview (cont'd)

## ■ Advantages

- ❑ can be varied to suit context
- ❑ issues can be explored more fully
- ❑ can elicit user views and identify unanticipated problems

## ■ Disadvantages

- ❑ very subjective
- ❑ time consuming

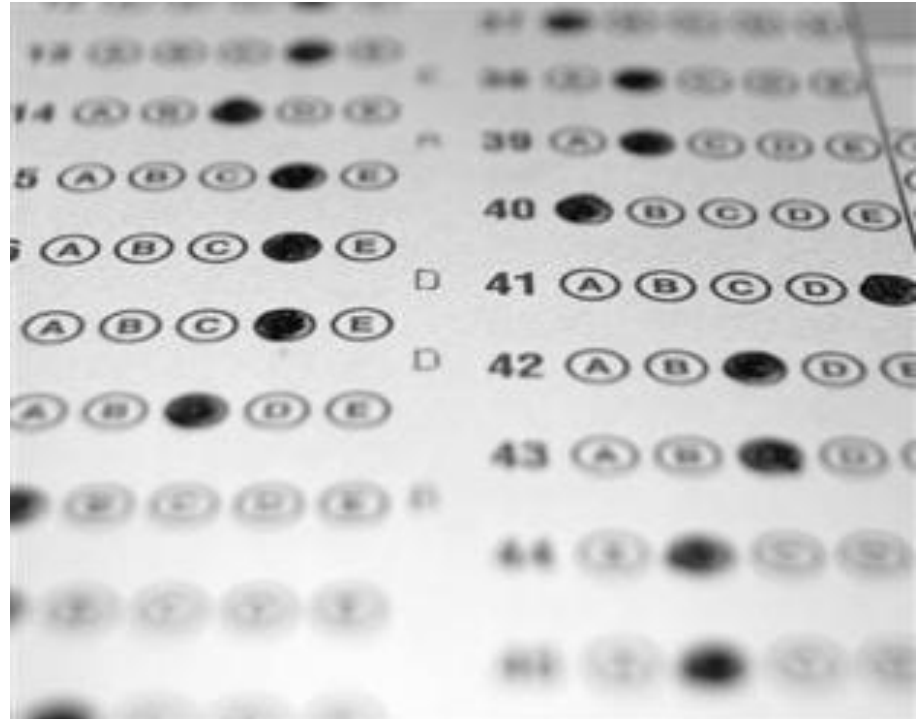
# Query techniques: Survey

- Prepare a set of fixed questions given to users
  - Need careful design
- Expensive preparation, but cheap administration
  - Use cheap methods: e.g. phone, mail, email, online form
  - Does not require presence of evaluator
- Results can be quantified

# Survey (cont'd)

## ■ Survey question types

- ❑ Open-ended
- ❑ Rating scalar
- ❑ Multi-choice
- ❑ Ranked



# Survey (cont'd)

## ■ Open-ended questions

- ❑ Asks for unprompted opinions
  - can you suggest any improvements to the systems?
- ❑ Don't restrict answers to evaluators' expectations
  
- ❑ Good for general subjective information
  - but difficult to analyze rigorously

# Survey (cont'd)

## ■ Rating scalar questions

- Ask user to judge a specific statement on a numeric scale

“The system is easy to recover from mistakes”

Disagree 1 2 3 4 5 Agree

- Scale usually corresponds with agreement or disagreement with a statement
  - typically 5 or 7 scale measuring agreement/disagreement with statements
- Easily to analysis, but needs careful design

# Survey (cont'd)

## ■ Multi-choice questions

- Respondent offered a choice of explicit responses

“Is there a clear way to return to a starting point or main menu?”

A. Yes, it's very clear

B. Maybe, but it's hard to navigate

C. No, I cannot find

# Survey (cont'd)

## ■ Ranked questions

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

Please rank the usefulness of these methods for text input on mobile device

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

\_\_\_ keypad

\_\_\_ QWERTY keyboard

\_\_\_ touch screen with virtual keyboard



# Survey (cont'd)

## ■ Advantages

- ❑ quick and reaches large user group
- ❑ does not require presence of evaluator
  - potential to automate administration and analysis
- ❑ can be analyzed more rigorously
  - results can be quantified

## ■ Disadvantages

- ❑ preparation can be expensive
- ❑ limited quality of feedback

## ■ Depends on having right questions

# Outline

- Overview of user interface design evaluation
- UI evaluation methods
  - Expert
    - Heuristic evaluation
  - User
    - Field studies
    - **Formative evaluation**
    - Controlled experiment



# Formative evaluation

- Basic steps similar to field studies
- But different
  - Users are assigned to certain tasks rather than natural
  - Evaluators
    - Choose participants
    - Assign each participant some tasks
    - Watch participants to perform tasks
- “Think aloud”

# Think aloud

- The goal is to know what the users are thinking, not just what they are doing
- First, ask the users to talk their thoughts while doing the task, example:
  - *what they are trying to do*
  - *why they took an action*
  - *how they interpret what the system did*
- Prompt the user to keep talking
  - – “tell me what you are thinking”
- Make a recording the thoughts
  - using note pad, audio or video tape

# Think aloud (cont'd)

- Widely used evaluation method in industry
- Advantages
  - simplicity - requires little expertise
  - can provide useful insight into the user's thinking
  - can show how system is actually use
- Disadvantages
  - subjective
  - selective
  - act of describing may alter task performance

# Example

- <http://www.youtube.com/watch?v=Aa-svs5mQD8&feature=related>

# Outline

- Overview of user interface design evaluation
- UI evaluation methods
  - Expert
    - Heuristic evaluation
  - User
    - Field studies
    - Formative evaluation
    - **Controlled experiment**



# Controlled experiment

- A preferred traditional scientific method
- Clear convincing result on specific issues
- Benchmark tests
- Allows system comparison, fine-tuning of details



# Controlled experiment (cont'd)

- Can deal with a practical problem and use the scientific method and theoretical framework
- Theory-driven, hypothesis testing
  - Study relations by manipulating **independent variables**
  - Observe effect on one or more **dependent variables**
    - Modify independent variables → measure dependent variables

# Controlled experiment (cont'd)

## ■ Procedure

- ❑ Prepare proposal
- ❑ Prepare environments
- ❑ Choose participants
- ❑ Determine forms and properties to be used
- ❑ Determine evaluation strategies

# 1. Prepare test proposal

- Test proposal
  - test objective
  - test participants
  - test environment and materials
  - test methodology to be used
    - detailed descriptions of systems/tasks/properties being tested
  - test measures
    - test data to be collected and analyzed

## 2. Prepare environments

- Usability testing labs
- Equipment
  - Computers
  - Video cameras
  - Internet access



- Video:

<http://www.youtube.com/watch?v=XtXYchttqQ4&feature=related>

# 3. Choose participants

- Participants should be chosen to represent the intended user communities, considering:
  - background, experience with the task, motivation, education, etc.
- Participation should always be voluntary and informed

# 4. Determine forms and properties

- Different forms
  - Paper mockups
  - Special usability testing
  - Remote usability testing
  - Etc.
- Videotaping participants performing tasks is often valuable

# 5. Evaluation strategies

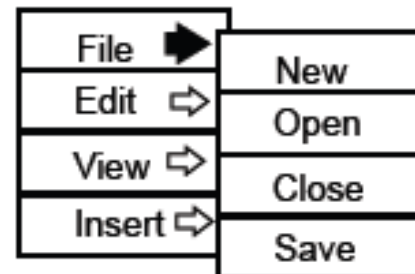
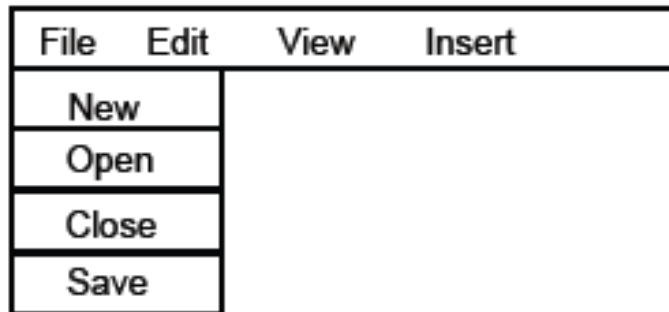
- The scientific method applied to HCI include major tasks:
  - ❑ A. state lucid and testable hypothesis
  - ❑ B. identify independent variables that are to be manipulated
  - ❑ C. choose dependent variables that will be measured
  - ❑ D. select subjects and assign subjects to groups
  - ❑ E. control for biasing factors
  - ❑ F. apply statistical methods to data analysis
  - ❑ G. interpret experimental results

# 5. Evaluation strategies (cont'd)

## ■ Hypothesis

- Example, testing menu

There is no difference in user performance (time and error rate) when selecting a single item from a pop-up or a pull down menu of 4 items, regardless of the subject's previous expertise in using a mouse or using the different menu types





# 5. Evaluation strategies (cont'd)

## ■ Independent variables

- things you control **independent** of a subject's behavior
- Example, in the menu testing example:

“There is no difference in user performance (time and error rate) when selecting a single item from a pop-up or a pull down menu of 4 items, regardless of the subject's previous expertise in using a mouse or using the different menu types”

### **Independent variables:**

- Menu type: pop-up and full-down
- Subject: expert and novice

# 5. Evaluation strategies (cont'd)

## ■ Dependent variables

- ❑ variables dependent on the subject' behavior / reaction to the independent variable
- ❑ the specific things you set out to quantitatively measure / observe
- ❑ Example, in the menu testing example:

### **Dependent variables:**

- Speed: time to select an item
- Error rate: # of errors made
- Easy to learn: time to learn to use menu thoroughly

# 5. Evaluation strategies (cont'd)

## ■ Subjects

- Those who perform tasks
- Carefully select ***subjects*** and assign subjects to groups

# 5. Evaluation strategies (cont'd)

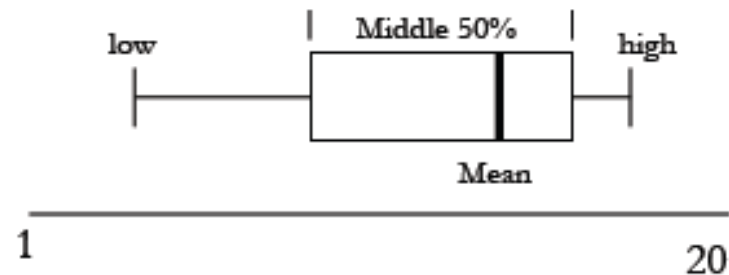
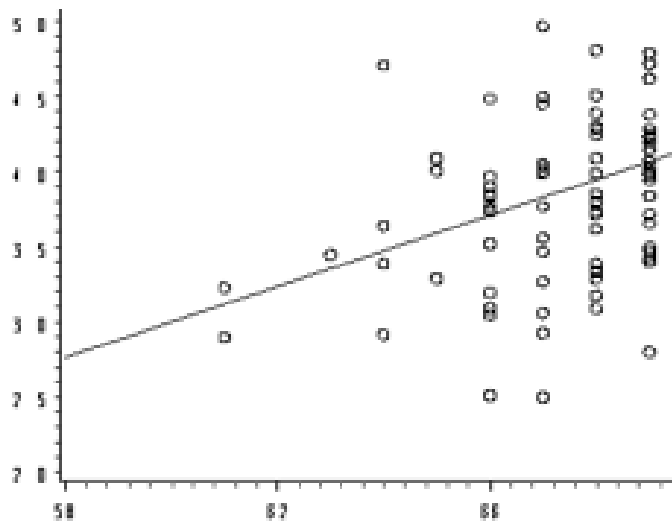
## ■ Statistical data analysis

- applying statistical methods to data analysis
- Simple analysis
  - Determine means and medians
  - Compare values of means and medians between methods

# 5. Evaluation strategies (cont'd)

## ■ Statistical data analysis

- Presentation using charts, tables, graphs



# 5. Evaluation strategies (cont'd)

## ■ Interpreting experimental results

- what you believe the results really mean
- their implications to your design
- how generalizable the results are
- limitations and critique

# Summary

- Overview of UI evaluation
  - Performed through the entire usability design lifecycle
- Field study (user testing) procedure
  - Preparing test proposal
  - Choosing participants
  - Selecting properties to be tested
  - Performing the tests. Techniques:
    - Inspection
    - Direct observation
    - Query techniques
  - Measuring the test results

# Summary (cont'd)

- Formative evaluation
  - Think aloud
- Controlled experiment
  - Major tasks of controlled experiments
    - Defining hypothesis
    - Determining independent and dependent variables
    - Selecting subjects
    - Performing statistical analysis
    - Interpreting results