1 USABILITY EVALUATION TECHNIQUES

Why, When, What and How to test

- Evaluation of the current design state is an important aspect of the User Centered Design, it can help answer the design team's question using real/representative users
- Why – it can ensure application can be used, it works as expected, meets particular criteria, and measure productivity gains between using this application and another
- What – part/all of an application, competing designs, icons/graphics, online help
- When – Prototype(paper/electronic), alpha/beta code, deployed application (different design might need different aspects of testing)

Test preparation

- Basic preparation
    - Set goals – can be Usability Goals (Learnability, efficiency …) or User experience goals – should be depends on the nature of the application and criteria
        - Usability goals should be measurable, objective and concrete, e.g. number of errors, duration to complete tasks…
    - determine what and how to measure
        - Initial Reactions, Users exploration, Successful completion of tasks, Timely completion of tasks, How well tasks are supported, User's satisfaction, Errors
        - The above tasks can be measured in: time required to learn/complete tasks, number of errors, and the severity of occurred errors, amount of assistant/help required, rating and comments
        - Test the ones are difficult to design/document/teach, and scientific method need to be considered
    - decide how to collect data
        - Verbal Protocols
            - Useful for establishing learnability and effectiveness
            - User think aloud while working on the scenario or tasks individually, facilitator might assist if user is frustrated
        - Active intervention 介入
            - Useful for establish effectiveness
            - Facilitator asks specific questions when users working on task, requires skill in knowing when to intervene
        - Co-discovery
            - Useful for assessing general
            - Having two or more participants, they can talk to each other. Can find people to know each other, let the one who's not dominant to drive
        - Questionnaire
            - Capture demographic 人口结构 information
            - Capture feedback on usefulness and usability of the application
            - Capture opinion not expressed during the test
            - Keep questions short and use plain language
    - Error Handling
        - Give user a hint after x min/errors
        - Give answer x minutes or move to the next task
    - Define scenarios and tasks
        - Task Scenario – Provides context for usability testing, it is a representation of actual work that a user would likely to perform to interact with a system. Do not use persons name and provide usability test participants with end goal and requirement. (not too detailed)
            - Put function into context, describe users goal and usage.
            - Include actions and artifacts used, make clear whether action is independent or part of another goal
            - Include special processing conditions

        - "Why the user is doing X"
    - Context Scenario – used in early in UCD lifecycle to help identify requirements, not focus on technology but design possibilities.
  - select participants
    - Select the representative user from each category wide range of experience/usage
    - Cross-section 有代表的实例 of user population, try to select randomly (scientific method)
    - Pick 4-5 per category
    - When trends emerge ⇔ stop, otherwise continue
    - Problem Found is dependent of number of test users, total number of usability problems, probability of finding problem with single users
  - define test roles
    - Test Co-Ordinator
    - Test Facilitator
    - Observers
    - Camera/Equipment operator
- Preparation for the test
  - Prepare the materials
    - Need a design to evaluate 需要有一个用来 evaluate 的设计例如 paper prototype
      - Make sure the design/prototype contains the functions for user to accomplish tasks
    - Consent form (ethical)
    - Pre/Post surveys & Questionnaire
    - Need logging sheets 用来记录用户的操作或者反馈, 和 template 类似
      - Task-specific log sheet 包含了正确的 path steps
      - General log sheet (with space for errors, accuracy and time)
  - Create a Usability Test Plan
    - 记录了用户需要 accomplish 的 tasks, documentation about the execution of the tests
      - Document the script for *the task scenarios* (give participant high level goal but doesn't describe how to do it)
    - might contain different aspects
    - What are the sources needed to run the test, and who is responsible for different aspects of evaluation
    - Contact details, consent froms
    - Designs that reflect the important paths through the design – important/common tasks people do
  - Invite and brief users and observers
  - Select environment and setup equipment
  - Select and obtain software and hardware
  - Conduct a dry run – identify issues with the process
    - Ensure the scenarios/tasks are able to perform in the current prototype/hardware/software
    - Ensure the tasks make sense
    - Make sure the questionnaire and logging sheet are validate
    - Provide estimate of time required for each participants

Running the Test

- Ensure equipment is functional
- Ensure materials are available
- Ensure observers have assembled
- Introduce user to test procedure
- Indicate that the software is under testing but not the participant
- Participant signs consent form
- Complete demographic pre-questionnaire
- Facilitator instructs participant from a script – tell the task scenario …
- Participant completes tasks
  - Log sheets – summarize
- Complete feedback questionnaire

Comp3511 week6

- Thank participant
- Debriefing
    - Immediately after each participant, allow turnaround of result 一般在完成每一个受访者之后进行讨论
    - Facilitates collation of observations 主持人和观察者进行比对
    - Provides opportunity to summarize -> it can help to categorize/prioritize and obtain concensus 一致的 on issues
- Issue table

Analysis of test data – After Test

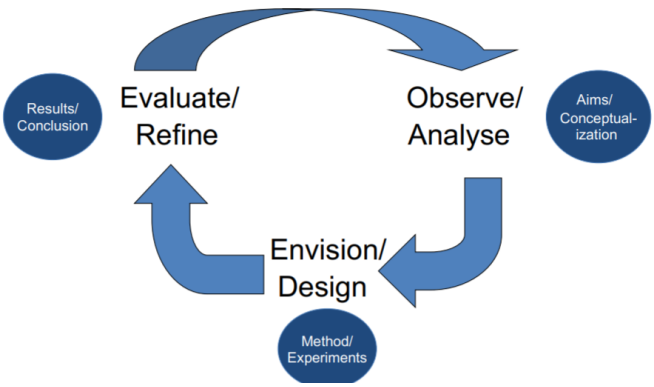
- Collate result – debriefing after the test can help the team to collation in order to document consensus result
- Interpret result
- Document result
    - Executive Summary – introduction – method – result and recommendations – appendices
- Deliver result
- Data Analysis
    - Summarize logging sheets
    - Get the result from debriefing section
    - Analyze the pre/post questionnaire result
    - Video
- Diagramming the Result – Affinity Diagramming
    - To identify the patterns in the observation and trends in analysis
- Prioritizing result – give severity ratings based on the frequency of problem occurs, impact if problem occurs, and the persistence of the problem

Other testing issues

- Usability testing labs 比较贵, 可能会 overkill, 可能会吓到用户并且需要用户 travel 到实验室
    - Usability & Testing areas
    - Audio/Video recording and editing equipment
    - Monitors for observing
    - Logging software
- Users work area 没有那么贵, 对于用户来说更加便捷, 但是移动 equipment 和 inappropriate/noisy 可能会影响
- Do-it-yourself testing
    - User install software
    - User step through tasks and record, responses on logging sheet and complete questionnaire
- Indirect observation
    - Diaries
    - Interaction logging
- Remote testing

2 SCIENTIFIC METHODOLOGY

In HCI context – evaluation of a new software application used with an organization is like a research.



Comp3511 week6

Experimental Research vs User Testing

- User Testing – aim to evaluate user's performance in order to improve the usability design
  - Need to measure the time to complete tasks and the number and type of errors
  - Conducted in real world environment
  - Few participant, usually is not replicable
  - Conditions controlled as much as possible
  - Procedure planned and result reported to developers
- Experimental Research – aim to discover new knowledge or answer a research question
  - Observe the relationship between two variables
  - Controlled environment, can be replicated
  - Need strongly controlled conditions, need experimental design
  - Scientific results reported to scientific community

Steps of the Scientific Method

- Observation
- Frames a Hypothesis – What are you expecting to find
- Methodology to test hypothesis – standardized
- Select Participants and collect data
- Describe, analyze, summarize data
- Discuss data and draw conclusion

Types of Research

- Laboratory experiments – High control, but artificial environment
- Quasi-Experiments – Partial control, no random assignments
- Non Experiments (Descriptive/correlational research) – Realistic setting, little control over extraneous effects

Some Useful Definitions

- Hypothesis – tentative 暂定的 belief about the relationship between two or more variables
- Variable – any phenomenon 现象 that can differ, or vary, from one situation to another or from one person to another. It is a characteristic that can take on different values
- Generalizability – Applicability of the findings to the entire population of interest

Important factors to consider

- Control extraneous fact – user's level of experience, Hawthrone Effect (depends on participants' awareness 受到额外的关注而绩效上升的情况), Ordering effect(实验的顺序可能会对受访者的表现或者结论造成影响)
- Control of biases – sampling, experimenter, double blind setup can be used (例如导师设置好实验设备, 另一个实验员并不知道什么正在被进行测试, 而受访者也不知道什么正在进行测试, 可以用来防止 demanding characteristics)
- Representative Sample (or random) – generalization from sample to population, sample size need to be considered
- Reliability – can the same result be obtained time and time again with same object
- Validity – are you measuring what you hope to measure
- Need a consistent and repeatable procedure
- Significance level – 你有多确定这个实验结果不是偶然?

Experimental Designs

- Different participants – single group of participants is allocated randomly to the experimental conditions 将类似的人任意的分到不同的实验环境里, 所以一个环境里可能会有来自很多不同组的人
  - No order effect
  - Many subject/individual are different -> cause problem
- Same participants – all participants appear in both conditions 将所有人不一定是一个组的人分到固定环境里, 所以在一个环境里都是同一个组的人
  - No individual differences

    - Counter-balancing needed because of ordering effect
- Matched participants – participants are matched in pairs, e.g. based on expertise/ gender …
    - No order effects, but individual differences reduced
    - Cannot be sure of perfect matching on all differences

3 COGNITIVE LOAD THEORY

Our Cognitive Architecture

- Huge long-term memory (LTM) used to store vast amounts of information over long periods of time, information stored in the form of schemas
- Limited working memory (WM) used to process current information, can only store a few items for a short period of time
- *Schema* allows us to bypass the limitation of Working Memory by Chunking large amounts of information together into single unit – allow us to ignore the huge amount of detail associated, relate information in long-term memory to fill the details have been perceived (2 most components of learning)
- *Automation* helps to reduce burden on working memory by allowing us to process information with minimal use of our limited working memory capacity
- *Automated Schema* – allow us to store information in long-term memory in efficient form + reduce the burden in wm

Worked out examples as an alternative learning strategy to means-ends analysis

- Means-ends analysis is useful problem solving technique but not a good learning strategy because it is really cognitive demanding focuses all the attention on retrieving a goal
- Worked out example or tutorial can help new users to learn (instead of through trial/error)

The split attention effect

- When information is structured to require people to mentally integrate information that are physically split, can cause unnecessary burden on our limited working memory
- Should make sure the information requires mentally integrate are physically integrate too, to reduce the strain 拉紧 on our limited working memory and frees up cognitive capacity for other processes

The redundancy Effect

- If information is not essential for understanding – is redundant, it is better to omit this information
- Processing non-essential information uses up valuable cognitive resources
- Screen should be designed only contain essential information, redundant text and graphics should be eliminated. E.g. google vs yahoo

Whether information is essential or redundant is depends on both nature of the materials and the level of expertise of users

Expertise reversal effect

- What is a good presentation format for novices learners, may not be an ideal format for experts

Reduce search

- Searching for related information on a screen is cognitively demanding, and wastes valuable cognitive resource
- Try to reduce search by:
    - Not putting too much information on single screen/page
    - Use consistent screen layout, so people can learn where to find things
    - Highlighting important information – color, font, bold
    - Making all web pages accessible from the home page

Diagrams

- If relationship between elements is complex, and diagram is meaningful, familiar, not too abstract ⇔ diagram can improve understanding
- Diagram can reduce the load on working memory because

- It can help to make the relationship between problem elements more concrete/explicit
- Help to reduce search for related information
- Act as external memory, help us to visualize the whole problem at once
- Help to represent complex relations between elements of problem in less abstract format
- Eliminate unnecessary information and focus on more essential things

The modality effect

- Research suggests that Working Memory has partially separate processors for handling visual and auditory information
- Presenting information in both way can increase the capacity of working memory.
- It works by expanding our effective working memory capacity
- Audio-visual materials are most useful as a method of eliminating the split attention effect, its also a useful way to present error messages
- Should avoid excessively large/complex auditory component, and timing of the audio is important

Animation/Transient information/gesture

- (Animation)
  - Animation can be distracting if they are redundant
  - They are transitory(so they're difficult to learn and overload working memory)
  - often need user control and interactivity
  - useful for more knowledgeable users
  - good for depicting human movement based tasks (例如老师手指黑板)
  - Better if more realistic
- (Transient Information effect)
  - Information that is transient should be used sparingly online, for learning
- (Gesture Research)
  - Adult
    - Useful to improve learning for static graphic format
    - Redundant/reduce learning for animation format, as human movement is already available in this format
  - For children, useful in both instructional format
  - So mimicking gesture is more useful for animation

Different sources of cognitive load

- Extrinsic/extraneous Cognitive Load
  - Exists as result of instructional design
  - Can be modified by reformatting the information
  - Want to reduce and keep minimum
- Intrinsic Cognition Load
  - Exists as a result of the type of material system is dealing with
  - Cannot be modifies
- Germane Cognitive Load
  - Exists as result of instructional design, associated with the effort to learn, result in resources being devoted to schema acquisition and learning
  - Want to strive to increase to improve learning