25th February 2020
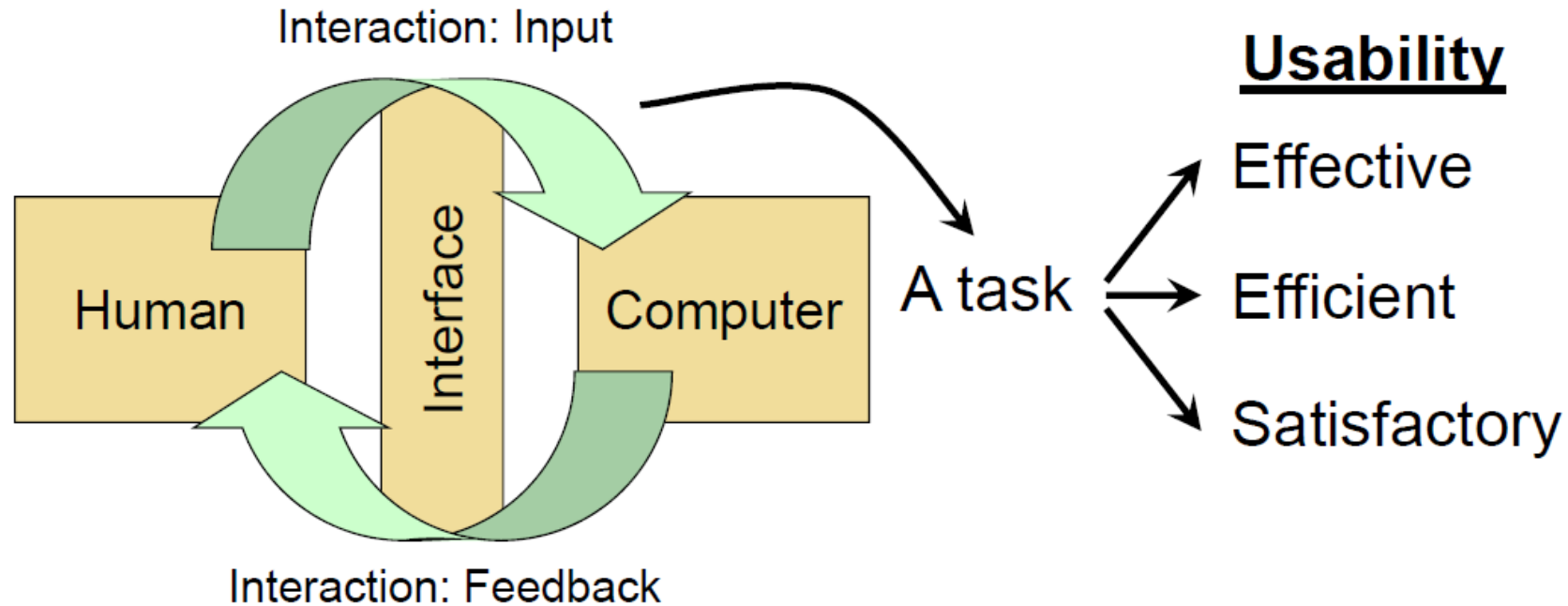# CSCM39/CSDM001: Human Computer Interaction

Dr Siyuan Liu

Department of Computer Science

siyuan.liu@swansea.ac.uk

Office Hour: Thursday 2-4pm

Usability testing

What is usability?

The official ISO 9241-11 definition of usability is: "the extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency* and *satisfaction* in a specified context of use."

What is usability testing?

- Representative users attempting representative tasks
- The world of usability testing includes:
  - Testing prototypes that have only been built on paper (known as paper prototypes);
  - Testing screen mock-ups or wireframes which have no functionality
  - Testing screen layouts which have partial functionality
  - Testing prototypes that look complete but have a human behind the scenes responding (known as the "Wizard of Oz" technique);
  - Testing working versions of software before it is officially released;
  - Testing software that has already been implemented in existing systems.

What is usability testing?

- One basic goal: to identify and fix interface-related flaws as early as possible
- What is an interface flaw
  - it's about interface components that impact on performance and frustration
  - It is not about general style preferences
- We are researching the interface, not the user
  - Discovering interface flaws that cause problems for a majority of people

/// **U N I T E D**

Worldwide sites | About United | Customer service   [   ]   [Search]   Fly **Ted**

Planning travel ▾   **Travel support ▾**   Mileage Plus ▾   Flights | Hotels | Cars | Vacations | Cruises | Deals

Home > Travel support > EasyCheck-in   ❓ Help

EasyCheck-in: Economy Plus offer

Log out

Would you like to purchase a seat in Economy Plus®?

Economy Plus offers more personal space and comfort with up to five more inches of legroom, as well as earlier boarding and deplaning.

This offer applies to all passengers in your reservation for the following flights only.

Flight/status	From	To
United 377 **Not departed** Boeing 757	**Baltimore, MD (BWI)** Nov 3 Scheduled: 12:52 PM Estimated: 12:52 PM	**Chicago, IL (ORD)** Nov 3 Scheduled: 2:02 PM Estimated: 2:02 PM

Price: 34.00 USD per person

Decline offer and continue   **Accept offer** ➡

Scope of usability testing

- Usability testing can involve, e.g.,
  - Having hundreds of users test interface, in different treatment groups
  - Sitting next to a total of 3 users and watching as they attempt different tasks
- The best usability testing is the one that actually take place
  - Be flexible
  - Be easier
  - Be quicker

Comparing with classical research methods

- A close cousin of traditional research methods
- Observation, key logging, click-stream analysis, quantitative measurement such as tasks and time performance, anonymity of participants, are all part of usability testing
- The end goals of usability testing and classical research are typically different
  - An industry-based approach to improving interfaces
  - Little concern for using only one approach or having strict controls
  - Building a successful product using the fewest resources, the fewest risks, in the shortest amount of time
- Many companies keep their usability testing results confidential, rather than publishing them as research

Types of usability testing

- Three general categories
  - Expert-based testing (usability inspections)
  - Automated testing
  - User-based testing
- All of these focus on usability, but when we talk about "usability testing" we generally are referring to representative users attempting representative tasks (user-based testing)

Expert-based testing

- An expert test is a structured inspection of an interface by one or more experts

- The people who developed the prototype interface being evaluated should not be involved with the expert review

- Expert-based tests should always come before user-based tests

- Different types of inspections, more common:
  - Heuristic review
  - Consistency inspections
  - Cognitive walkthrough

Shneiderman's 8 Golden Rules of Interface Design

- Strive for consistency
- Cater to universal usability
- Offer informative feedback
- Design dialogs to yield closure
- Prevent errors
- Permit easy reversal of actions
- Support internal locus of control
- Reduce short-term memory load

Expert-based testing

- Different types of inspections
  - Heuristic review
  - Consistency inspections
    - One or more experts review a series of screens or web pages for issues of consistencies
  - Cognitive walkthrough
    - Experts simulate users, "waling through" a series of tasks
  - Guideline review,
    - Expert compares a set of interfaces to a previously written set of interface guidelines, e.g., Web Content Accessibility Guidelines (WCAG, http://www.w3.org/WAI).
    - A large set of guidelines (10-200)
  - Others: formal usability inspection, pluralistic walkthrough

Automated usability testing

- Software applications, or web-based applications, examine a series of interfaces, and compare them to a known set of interface guidelines

- Often used when a large number of interfaces need to be examined and little time is available to do human-based reviews

- Automated tools often require manual checks

- Automated tools are often better at determining the presence of interface features, rather than whether those features are implemented properly

User-based testing

- A set of representative users attempting to perform a set of representative tasks

- User-based testing would take place during all stages of development
  - The earlier that it can take place, the better, when the results can influence the design more and when costs to make changes are much lower.

- We need usability testing because designers are not users and even the users themselves often have trouble identifying their interface needs

Types of user-based testing

- Formative testing
  - Takes place at an early stage of development, may include wireframes, paper, or other low-fidelity prototypes
  - More communication between test moderators and participants
  - More of a focus on how the user perceives an interface component, rather than how the user completes a tasks
  - Low-fidelity prototypes are low-cost, can be quickly developed, and multiple options tested

Types of user-based testing

- Summative testing
  - Takes place when there is a more formal prototype (high-fidelity prototypes) ready
  - The goal is to evaluate the effectiveness of design choices
  - Users may be more hesitant to criticize the interface at this point, since they may feel that it won't matter

Types of user-based testing

- Validation testing
  - Takes place right before an interface is released to the general user population
  - Compared to a set of benchmarks for other interfaces
  - The goal is to ensure, e.g.,
    - Can 90% of users complete specific tasks within 1 minute?
    - Repeated changes may be made until the specific task/time performance goals are reached
  - Occurs much less frequently than formative or summative testing

Stages of usability testing

1. Select and recruit representative users
2. Select the setting (home, workplace, usability lab, etc.)
3. Decide what tasks users should perform
4. Decide what type of data to collect and how to measure
5. Planning before the test session
6. Actions during the test session
7. Analysis after the test session

How many users are sufficient?

- It's still a hotly debated question
- Often, people will say, "5 users" even though there is no research basis for this, and it's misleading; recent research has found that 10 or perhaps more are needed
- You don't know in advance how many users you need, because it depends on how complex an interface is, and how many interface flaws exist
- The answer: get as many users as you afford to have, given the limitations of project budget, timeline, and organizational politics

Locations for usability testing

- A fixed usability laboratory
  - Two separate rooms, one for the participant, one for the observer, separated by a one-way mirror, with a video/audio feed
- In the user's workplace or home
  - Observe them by sitting near them, or using portable video equipment
- Remote usability testing, where you observe users over the web or through videoconferencing

Source: Photo by Elizabeth Buie for UserWorks, Inc., a usability consulting firm located in Silver Spring, MD, www.userworks.com.

Benefits and drawbacks of remote usability testing

Benefit	Drawback
Easy access to a greater number of participants	Difficult or impossible to pick up nonverbal and interpersonal cues
Participants have more flexibility in participating in a usability test on their own schedule, and researchers can run multiple usability tests at the same time	Hard (or impossible) to provide instructions when things "go wrong"
Easy collection and analysis of clickstream data	Researchers cannot ask any probing questions based on what occurs
Works better for summative testing, when your are collecting quantitative metrics	Researchers often miss the context of what was happening

Task list

- Unless the usability testing is very early stage and formative where you are only looking for impressions on general design choices, you will need to create a task list

- Task lists make users goal-directed

- Tasks need to be clear, unambiguous, and not require further explanation

- Make sure tasks can only be completed by using the interface (i.e. not something that is commonly known)

- Choose tasks that are critical, performed often and central to user goals

- Do not require the use of user's personal data

Task list

- Be clear in your instructions:
  - Must the tasks be completed in a certain order? Can users skip around?
  - Do certain tasks require that others are completed first?
  - How should participants note if they are giving up on a task and moving on?
  - Is there a time limit for the entire session?
  - Are there any test (dummy) accounts or fake identities that will need to be utilized as a part of the usability testing?
- Whether interventions will be allowed?

Measurement

- Often, you are interested in quantitative measurements
  - How long does it take for typical users to learn relevant tasks? (Time to learn)
  - How long does it take to perform relevant benchmarks? (Speed of performance)
  - How many and what kinds of errors are made during benchmark tasks? (Rate of errors by users)
  - How well do users maintain their knowledge after an hour, a day or a week? Frequency of use and ease of learning help make for better user retention (Retention over time)
  - How much did users like using various aspects of the interface? Allow for user feedback via interviews, free-form comments and satisfaction scales (Subjective satisfaction)
- Make sure to decide in advance how to collect this data
- In usability testing, especially formative usability testing, qualitative data is often as important as quantitative data

Different types of applications

- Life-critical systems
- Industrial and commercial uses
- Home and entertainment
- Exploratory, creative, collaborative applications
- Social/Technological applications

Different applications have different preferences for these measurements

Life-critical systems

- Applications:
  - Air traffic control
  - Nuclear reactors
  - Power utilities
  - Police & fire dispatch systems
- Requirements:
  - Reliability and effectiveness
  - Error-free performance
- Not as important
  - Cost
  - Training time
  - User satisfaction: users are well-motived professionals

Industrial and commercial uses

- Applications
  - Banking
  - Insurance
  - Order entry
  - Inventory management
  - Reservation
  - Billing system

- Requirements
  - Ease of learning is important to reduce training costs
  - Speed and error rates are relative to cost
  - Subjective satisfaction is fairly important to limit operator burnout
  - Speed of performance is important because of the number of transactions

Office, home and entertainment

- Applications:
  - Personal computing App
  - Word processing, electronic mail, computer conferencing, and video game system, educational packages, search engines, mobile device, etc.
- Requirements:
  - Ease of learning, low error rates, and subjective satisfaction are paramount due to use is often discretionary and competition fierce
  - Interfaces must be intuitive with easy –to-use online help
  - Market competition often forces the need for low cost
  - Choosing functionality is difficult because the population has a wide rang of both novice and expert users
    - A layered or level-structures design is an approach to graceful evolution, i.e., the design of search engines, which almost always have a basic and advanced interface.

Exploratory, creative & collaborative systems

- Applications:
  - For supporting human creativity
  - Web browsing, search engines, artist toolkits, architectural design, software development, music composition, and scientific modelling systems
- Requirements:
  - Users may be knowledgeable in the task domain but novices in the underlying computer concepts (however, high in their expectations on the usability), so
    - Due to exploratory nature, these systems are hard to design and evaluate
    - Designers should pursue the goal: "computer/UI should vanish" so that the user
    - can be absorbed in their task domains

Social/Technical systems

- Applications (usually used by government)
  - Voting, health support, identity verification, crime reporting
- Requirements:
  - Trust, privacy, responsibility, and security are issues
  - Verifiable sources and status feedback are important
  - Diverse levels of users: Ease of learning for novices and feedback to build trust
  - Administrators need tools to detect unusual patterns of usage

The testing session

- Confirm the time and location with participants beforehand
- Make sure to leave extra time in your schedule
- Test any equipment that you will need
- Participants must be informed of their rights and sign an informed consent form and confidential agreement if necessary
- Clarify how participants will be paid for their time
- Make sure to remind participants that they are not being tested
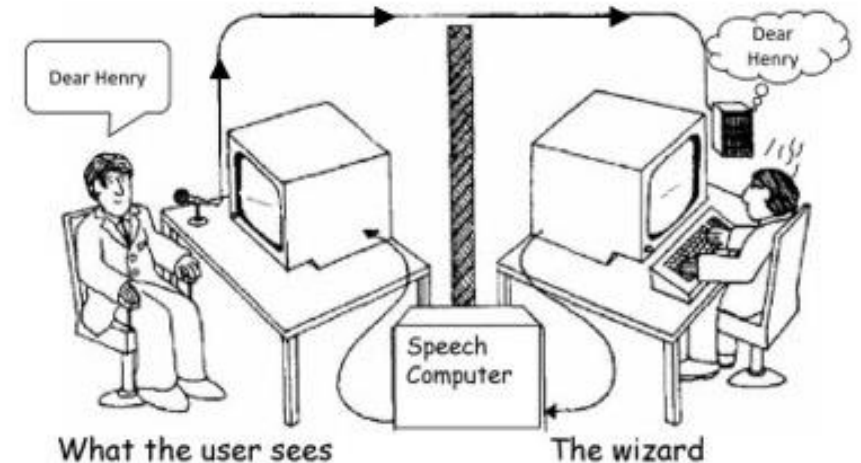- Remind participants that their feedback is important!

Making sense of the data

- Often, inferential statistics are not the goal; just descriptive statistics, and "telling the story"
- Write up a report, short, clear, to the point, so that people will actually read it
- The goal is to have the testing results influence the interface design
- Consider carefully who will be reading the report, e.g., interface designers, software engineers, project managers, and other managers involved in software development
- Not all flaws will be fixed…so, summarize and prioritize which are most important!
- Suggest fixes for each flaw
- Again, never include data that identifies the participants

Other variations of usability testing

- A/B testing
  - Minor tweaks are made to interfaces (e.g. websites) that are already in real-world use and data is collected
- Technology probes
  - Place a new technology into a real-world setting to see how it's used and how people react and interact
  - Sort of a hybrid of usability testing and ethnography
- Wizard-of-oz method
  - A non-functional prototype is used for user-based testing, but a human is providing the responses to the user, as if it was the computer providing the responses

Wizard of Oz testing – The listening type writer IBM 1984

Dear Henry

Dear Henry

Speech Computer

What the user sees

The wizard

Wizard of Oz testing – The listening typewriter IBM 1984