CSCC10

Human-Computer Interaction

Lecture 8: Data Analysis & Experimental Design

Dr. Naureen Nizam

nizam@cs.toronto.edu

July 3, 2024

Administrivia

- Project Phase III
  - Due: **July 9, 2024**
- **Tutorial:**
  - This Week – Reading
- Assignment 2
  - Due: **July 7, 2024 (New Date)**

Final Exam (Tentative)

Monday, August 19, 2024
2:00 – 4:00 PM (EST)
In-Person
IC 220

The plan for today …

- Data Analysis & Experimental Design

Overview

- Qualitative Data Analysis
- Quantitative Data Analysis
- How to analyze data gathered from:
  - Questionnaires
  - Interviews
  - Observations
- Software Packages that help you analyze?
- How to present findings?

Qualitative & Quantitative Data

- **Quantitative data**
  - expressed as numbers
  - numerical methods to ascertain size, magnitude, amount
- **Qualitative data**
  - difficult to measure sensibly as numbers, e.g. count number of words to measure dissatisfaction
  - expresses the nature of elements and is represented as themes, patterns, stories
- **Be careful how you manipulate data and numbers!**

Question!

Quantitative Analysis

An average person is 5ft. 11 inches tall, weights 180 pounds, and is 45 years old.

Qualitative Analysis

An average person is tall, thin, and middle-aged.

7

Qualitative & Quantitative Data

Source: http://valientmarketresearch.com/tag/qualitative/

8

Question!

Certain forms of data gathering can only result in quantitative data and others can only result in qualitative data.

True | False

Questionnaires?

Observations?

10

Data Gathered and Initial Steps

	Raw Data Format	Example qualitative Data	Example quantitative Data	Initital processing steps
Interviews	• Audio recordings • Video recordings • Interviewee notes.	• Resp. to open ended questions. • Video pics • Opinions	Age, job, role, years of exp. Resp. to closed questions.	• Transcription of recording • Expansion of notes.
Questionnaires	• Written responses • Online database	• Resp. to open ended questions. • "Comment" fields • Opinions	Age, job, role, years of exp., Resp. to closed question,	• Clean up data • Filter into different data sets.
Observation	• Observer's notes • Photographs • Audio/Video recordings • Data Logs • Think-aloud notes	• Records of behavior • Copies of informal proc. • Task desc.	Demographics of participants, time spent on task, # of participants	• Expansion of notes. • Transcription of recordings • Sync. Btw data and recordings.

11

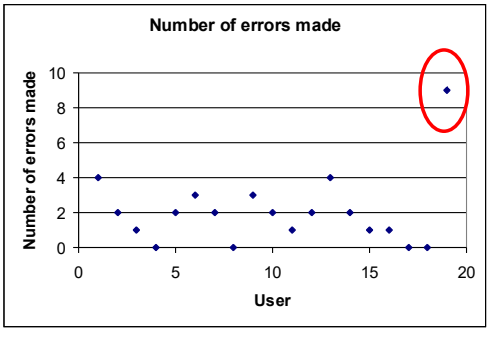
Quantitative Analysis

- Averages
  - Mean: add up values and divide by number of data points
  - Median: middle value of data when ranked
  - Mode: figure that appears most often in the data

  Example: {2, 3, 4, 6, 6, 7, 7, 7, 8}
- Percentages
- 50% (1/2)
- Be careful not to mislead with numbers!
- Graphical representations give overview of data

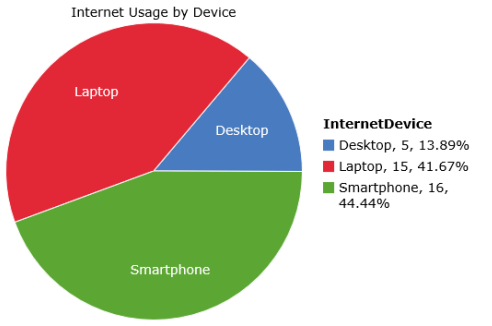
12

Quantitative Analysis



13

Quantitative Analysis



Quantitative Analysis

- Visualizing Log Data – interaction profiles of players in online game.

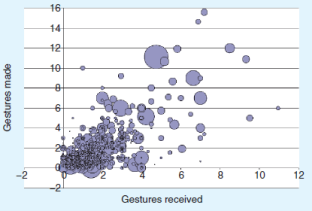
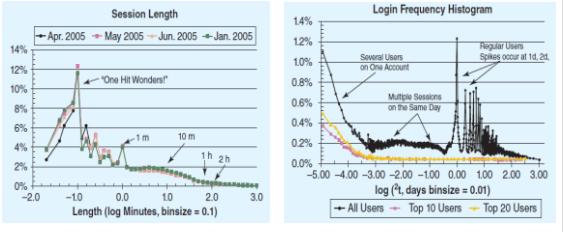


Figure 8.6 Interaction profiles of players in the cantina

Source: N. Ducheneaut and R.J. Morris (2004): "The social side of gaming: a study of interaction patterns in a massively multiplayer online game" in *Proceedings of CSCW 04*. ©2004 Association for Computing Machinery, Inc. Reprinted by permission.

Web Analytics



session length data of four different months from Teachers' Domain (NSDL)

Source: Khoo, M., Pagano, J., Washington, A. L., Recker, M., Palmer, B., and Donahue, R. A. (2008) Using web metrics to analyze digital libraries. *Proceedings of Joint Conference on Digital Libraries*, Pittsburgh, June 16–20. ©2008 Association for Computing Machinery, Inc. Reprinted by permission.

CHI 2018 - Proceedings

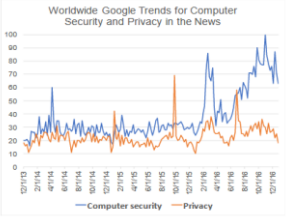


Figure 1. Worldwide Google Trends for the "computer security" and "privacy" topics from December 1st, 2013 to December 31st, 2016. The two topics have steadily been gaining increasing media exposure.

Source: Breaking! A Typology of Security and Privacy News and How It's Shared
Sauvik Das, *Georgia Institute of Technology,* Joanne Lo, *Carnegie Mellon University*
Laura Dabbish, *Carnegie Mellon University,* Jason I. Hong, *Carnegie Mellon University*

CHI 2018 - Proceedings

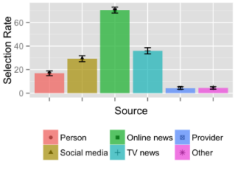


Figure 2. Distribution of how participants reported hearing about security and privacy news events. Most people heard about news events through online news sources.

Source: Breaking! A Typology of Security and Privacy News and How It's Shared
Sauvik Das, *Georgia Institute of Technology,* Joanne Lo, *Carnegie Mellon University*
Laura Dabbish, *Carnegie Mellon University,* Jason I. Hong, *Carnegie Mellon University*

CHI 2018 - Proceedings

	Online News	Another Person	Social Media	TV / Video	Service Provider
Intercept	0.93˚	-0.97˚	-0.25	-1.28˚	-3.01˚
Individual-level variables					
Age	-0.10*	-0.20˚	-0.24*	0.21˚	0.13
Male (vs. Female)	0.35†	-0.21	-0.04	-0.13	-0.89†
Security Behavioral Intention	0.38˚	-0.11	-0.11	-0.12	0.04
Event-type comparisons					
F vs. C	0.42	0.29	0.36	0.17	0.41
P vs. C	0.02	0.04	0.20	-0.22	-0.79*
S vs. C	-0.12	0.27	-0.40	0.01	-0.08
P vs. F	-0.41	-0.25	-0.16	-0.39	-1.40†
S vs. F	-0.54	-0.02	-0.76*	-0.16	-0.49
S vs. P	-0.14	0.23	-0.60*	0.23	0.70

Table 2. Logistic regression coefficients of information source modeled against individual-level factors and event-type comparisons. Rows represent IVs, columns DVs. Both individual-level factors (age, gender, security behavioral intention) and event-types significantly correlated with how people heard about news events.

*˚p<0.001 † p <0.01 * p<0.05*

F=financial data breaches, C=corporate data breaches, S=high sensitivity systems breaches, P=politicized / activist cybersecurity

Source: Breaking! A Typology of Security and Privacy News and How It's Shared
Sauvik Das, *Georgia Institute of Technology,* Joanne Lo, *Carnegie Mellon University*
Laura Dabbish, *Carnegie Mellon University,* Jason I. Hong, *Carnegie Mellon University*

ID	Age/ Sex	Diagnosis C: Congenital, A: Adventitious	Visual Acuity L: Left, R: Right	Tools Used ZT: ZoomText, VO: VoiceOver, OM: Optelec Magnifier
P1	35/M	Retinitis pigmentosa (C)	L:20/200, R:0	OM, MagicPro, JAWS, ZT
P2	48/M	Macular Telangiectasia (A)	20/150	Windows Magnifier, ZoomText
P3	46/M	Congenital cataracts (C)	L:20/240, R:20/200	OM, Pocket magnifier, Audio book, NVDA, ZT
P4	64/F	Retinopathy of prematurity (C)	L:20/200, R:20/200	Hand magnifier, Telescopic lens, iPhone camera, ZT
P5	62/F	Macular degeneration (C)	L:0, R:20/300	CCTV, Window 's Magnifier, ZT
P6	37/M	Congenital cataracts (C)	20/800	JAWS, NVDA, VO, Mac's Zoom, ZT
P7	34/M	Albinism (C)	L:20/200, R:20/400	NVDA, AppVision, GW-Micro, Large display, iPhone, ZT
P8	27/F	Myopia strabismus (C)	20/600	Magnifier, Narrator, iPhone camera, ZT
P9	29/F	Albinism (C)	20/240	OM, CCTV, Portable CCTV, Zoom, ZT
P10	70/M	Glaucoma (A)	Unknown	Magnifier, Narrator, Larger Key Caps, Telescopic lens, ZT
P11	33/F	ROP & Glaucoma (C)	L:20/200, R:20/400	Zoom, VO, JAWS, Handheld magnifier, ZT
P12	52/M	Optic atrophy (C)	20/800	JAWS, iPhone, ZT
P13	32/M	Nystagmus (C)	20/120	Telescopic lens, Magnifier, ZT
P14	26/F	Pathological Myopia (A)	20/200	JAWS, Magnifier, Phone camera, Large display, ZT
P15	31/F	Pathological Myopia (A)	20/280	Large display, Narrator, JAWS, Amazon Echo, ZT

Table 2. Demographic Information of the 15 Participants.

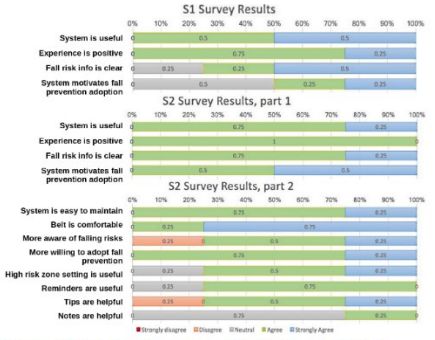


Figure 4: The results of the early- and post-deployment surveys.

Question \ Response (1 to 5)	1	2	3	4	5
1. How easy is it to rotate the Dial compared to panning with the mouse?	0	0	2	8	5
2. How easy is it to learn SteeringWheel?	1	2	0	4	8
3. How important is it to view the segment boundaries?	0	0	4	3	7
4. How useful is audio feedback?	0	1	6	6	2
5. How useful is haptic feedback?	0	0	0	6	9
6. How easy is it to perform gestures?	1	1	6	5	2
7. How easy is it to use Dial and Mouse together with both hands?	1	2	2	3	7
8. How easy is it to fill forms with SteeringWheel?	0	0	1	7	7
9. How easy is it to customize the interface with the Dashboard?	0	1	2	7	5
10. How noticeable were the effects of locality preservation?	0	0	1	9	5

Color Legends (5 color bins):

0-3 | 4-6 | 7-9 | 10-12 | 13-15

Table 3. Post-experiment questionnaire and responses. The columns labeled as 1 to 5 show how many times a particular

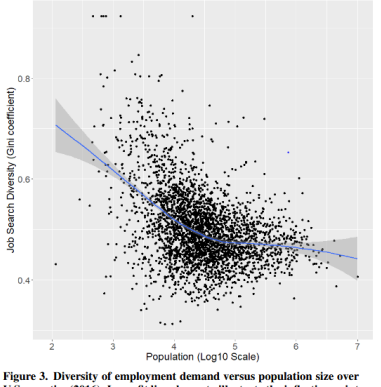


Figure 3. Diversity of employment demand versus population size over U.S. counties (2016). Loess fit line shown to illustrate the inflection point in job search diversity at population size of about 50,000.

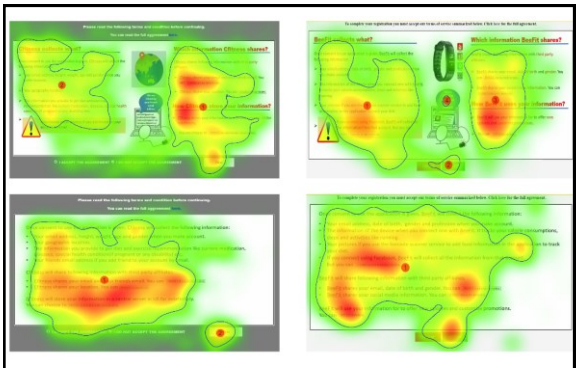


Figure 2. Aggregated heatmaps of eye gaze fixations. Row 1: Comic ToS (Interface 1: 18 participants, Interface 2: 16 participants). Row 2: Textured ToS (Interface 1: 18 participants, Interface 2: 17 participants). Row 3: Textured ToS (Interface 1: 17 participants, Interface 2: 20 participants)

Qualitative Analysis

- **Recurring patterns or themes**
  - Emergent from data, dependent on observation framework if used
- **Categorizing data**
  - Categorization scheme may be emergent or pre-specified
- **Looking for critical incidents**
  - Helps to focus in on key events

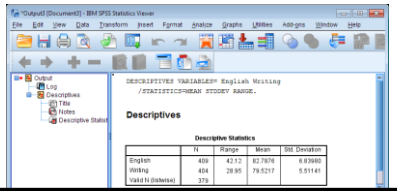
Figure 8.8 Building the affinity diagram of Indian ATM usage

Source: Figure 1, A. DeAngeli, U. Athavamker, A. Joshi, L. Coventry and G.I. Johnson (2004) "Introducing ATMs in India: a contextual inquiry", *Interacting with Computers* 16(1), 29–44. Reproduced with permission.

Tools to Support Data Analysis

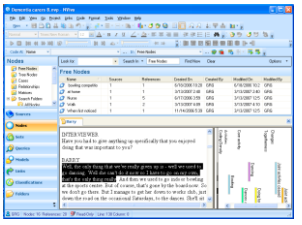
- Spreadsheet – simple to use, basic graphs (Microsoft Excel)

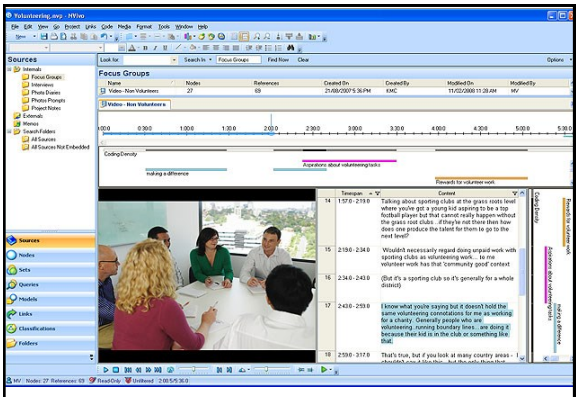
- Statistical packages, e.g. SPSS

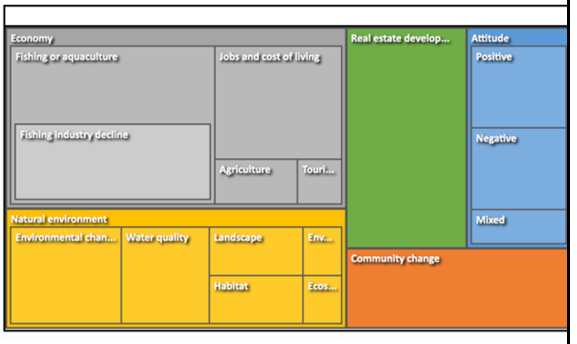


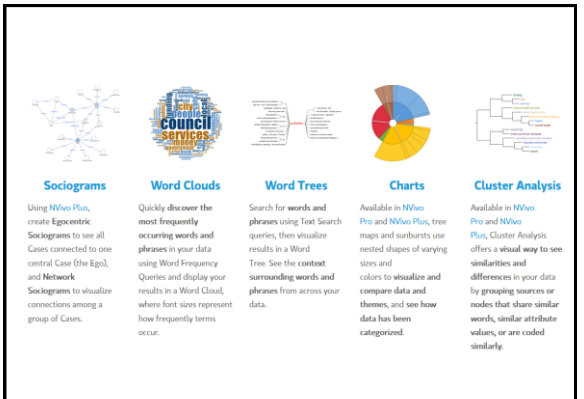
Tools to Support Data Analysis

- Qualitative data analysis tools
  - Categorization and theme-based analysis
  - Quantitative analysis of text-based data
    - Nvivo and Atlas.ti support qualitative data analysis









Theoretical Framework for Qualitative Analysis

- Basing data analysis around theoretical frameworks provides further insight
- Three such frameworks are:
  - Grounded Theory
  - Distributed Cognition
  - Activity Theory

Grounded Theory

- Aims to derive theory from systematic analysis of data
- Based on categorization approach (called here 'coding')
- Three levels of 'coding'
  - Open: identify categories (word, phrase, sentences, etc.)
  - Axial: flesh out and link to subcategories
  - Selective: form theoretical scheme (backbone of a theory)
- Researchers are encouraged to draw on own theoretical backgrounds to inform analysis

Grounded Theory: code book

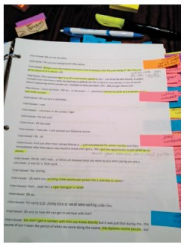


Figure 8.13 Code book used in a grounded theory analysis of citizens' motivations to contribute to citizen science

Source: Rotman, D. *et al* (2014). Does motivation in citizen science change with time and culture? In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW Companion '14).* ACM, New York, NY, USA, 229–232. ©2014 Association for Computing Machinery, Inc. Reprinted by permission.

Grounded Theory: axial coding

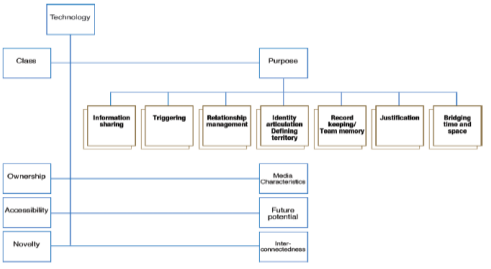


Figure 8.14 Axial coding for the technology category

Source: S. Sarker, F. Lau and S. Sahay (2001): "Using an adapted grounded theory approach for inductivetheory building about virtual team development". *The Data Base for Advances in Information Systems*, 32(1), pp. 38–56 ©2001 Association for Computing Machinery, Inc. Reprinted by permission.

Grounded Theory

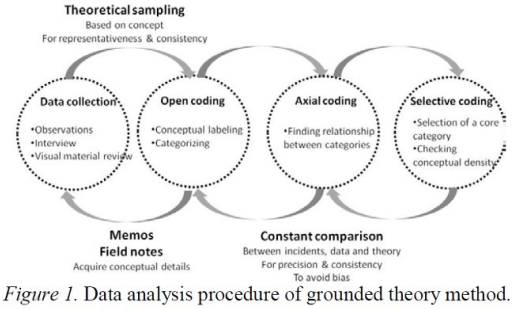


Figure 1. Data analysis procedure of grounded theory method.

Sourrce: http://dohwan.tistory.com/entry/Grounded-Theory%EC%99%80-Qualitative-Content-

Distributed Cognition

- The people, environment & artefacts are regarded as one cognitive system
- Used for analyzing collaborative work
- Focuses on information propagation & transformation

Activity Theory

- Explains human behaviour in terms of our practical activity in the world
- Provides a framework that focuses analysis around the concept of an 'activity' and helps to identify tensions between the different elements of the system

Theoretical Framework for Qualitative Analysis

- Basing data analysis around theoretical frameworks provides further insight
- Three such frameworks are:
  - Grounded Theory
  - Distributed Cognition
  - Activity Theory

Presenting the Findings

- Only make claims that your data can support*
- The best way to present your findings depends on the audience, the purpose, and the data gathering and analysis undertaken
- Graphical representations may be appropriate for presentation
- Other techniques are:
  - Rigorous notations, e.g. UML
  - Using stories, e.g. to create scenarios
  - Summarizing the findings

Exercise 1/3

- **Finding:** 2 out of 4 people who filled the questionnaire ticked the box that said they prefer not to use the ring-back functionality on their cell phone.
- **Statement:** Half of the users don't use the ring-back functionality.

Exercise 2/3

- **Finding:** John who works in the design department was observed one day walking for 10 minutes to collect printout from the high-quality colour printer.
- **Statement:** Significant time is wasted by designers who have to walk a long distance to collect printout.

Exercise 3/3

- **Finding:** A data log of 1000 hours of interaction with a website recorded during January, February and March records 8 hours spent looking at the help files/documents.
- **Statement:** The website's help files were used less than 1% of the time during the first quarter of the year.

Summary

- The data analysis that can be done depends on the data gathering that was done
- Qualitative and quantitative data may be gathered from any of the three main data gathering approaches
- Percentages and averages are commonly used in Interaction Design
- Mean, median and mode are different kinds of 'average' and can have very different answers for the same set of data
- Grounded Theory, Distributed Cognition and Activity Theory are theoretical frameworks to support data analysis
- Presentation of the findings should not overstate the evidence

Coding Exercise

Participant A:

We have an upstream delivery method, primarily to retailers, a 3rd party vendor keeps track and sends the bulbs out. The idea is to lower the price of shelves. The major constraint is it's difficult o know our customers – hard to evaluate who is buying the bulb and what sockets they are putting them into. However there are some advantages. We can control consumer choice, working with retail partners, we have dominant displays of incentivized bulbs. Yet we think EISA is going to really impact future savings….

Coding Exercise

Participant	Excerpt	Code	Themes or Categories
A			
A			
A			
A			
A			
A			

Coding Exercise

Participant A:

We have an upstream delivery method, primarily to retailers, a 3rd party vendor keeps track and sends the bulbs out. The idea is to lower the price of shelves. The major constraint is it's difficult o know our customers – hard to evaluate who is buying the bulb and what sockets they are putting them into. However there are some advantages. We can control consumer choice, working with retail partners, we have dominant displays of incentivized bulbs. Yet we think EISA is going to really impact future savings….

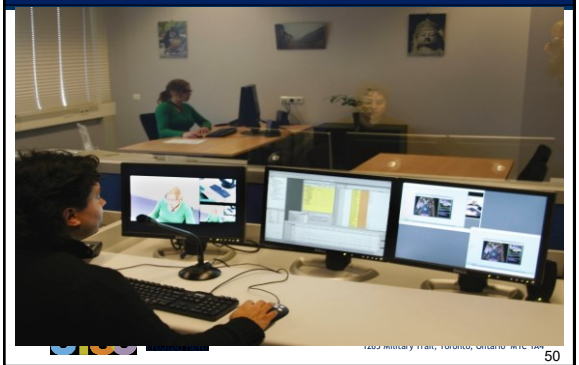
Coding Exercise

Participant	Excerpt	Code	Themes or Categories
1	Upstream Delivery Method	UP	Type of Program
1	Difficult to know our customers	NEG	Weaknesses of Program
1	Control Customer Choice	POS	Strengths of Program
1	EISA is going to really impact future savings	EISA	Impact of EISA

Experimental Design

- Usability Testing
- Experiments
- Field Studies

Recall



50

Usability Testing

- Involves recording performance of typical users doing typical tasks.
- Controlled settings.
- Users are observed and timed.
- Data is recorded on video & key presses are logged.
- The data is used to calculate performance times, and to identify & explain errors.
- User satisfaction is evaluated using questionnaires & interviews.
- Field observations may be used to provide contextual understanding.

51

Testing Conditions

- Usability lab or other controlled space.
- Emphasis on:
  - selecting representative users;
  - developing representative tasks.
- 5-10 users typically selected.
- Tasks usually around 30 minutes
- Test conditions are the same for every participant.
- Informed consent form explains procedures and deals with ethical issues.

52

Types of Data

- Time to complete a task.
- Time to complete a task after a specified time away from the product.
- Number and type of errors per task.
- Number of errors per unit of time.
- Number of times online help and manuals accessed.
- Number of users making an error.
- Number of users successfully completing a task.
- Number of clicks

53

Usability Testing: The iPad

- 7 participants with 3+ months experience with iPhones
- Signed an informed consent form explaining:
  - what the participant would be asked to do;
  - the length of time needed for the study;
  - the compensation that would be offered for participating;
  - participants' right to withdraw from the study at any time;
  - a promise that the person's identity would not be disclosed; and
  - an agreement that the data collected would be confidential and would be available to only the evaluators
- Then they were asked to explore the iPad
- Next they were asked to perform randomly assigned specified tasks

54

Examples of the Tasks

App or website	Task
iBook	Download a free copy of *Alice's Adventures in Wonderland* and read through the first few pages.
Craigslist	Find some free mulch for your garden.
eBay	You want to buy a new iPad on eBay. Find one that you could buy from a reputable seller.
Time Magazine	Browse through the magazine and find the best pictures of the week.
Epicurious	You want to make an apple pie for tonight. Find a recipe and see what you need to buy in order to prepare it.
Kayak	You are planning a trip to Death Valley in May this year. Find a hotel located in the park or close to the park.

Table 14.1 Examples of some of the tests used in the iPad evaluation (adapted from Budiu and Nielsen, 2010).

Source: Copyright Nielsen Norman Group, from report available at http://www.nngroup.com/reports/.

55

Example of the equipment

Figure 14.6 The setup used in the Chicago usability testing sessions
Source: Copyright Nielsen Norman Group, from report available at http://www.nngroup.com/reports/.

Problems & Actions

- Problems detected:
  - Accessing the Web was difficult
  - Lack of affordance and feedback
  - Getting lost
  - Knowing where to tap
- Actions by evaluators:
  - Reported to developers
  - Made available to public on nngroup.com
- Accessibility for all users important

Discussion Questions

1. Was the selection of participants for the iPad study appropriate?

2. What might have been the problems with asking participants to think out aloud as they complete the task.

Experiments

- Test hypothesis
- Predict the relationship between two or more variables.
- Independent variable is manipulated by the researcher.
- Dependent variable influenced by the independent variable.
- Typical experimental designs have one or two independent variables.
- Validated statistically & replicable.

Experimental Designs

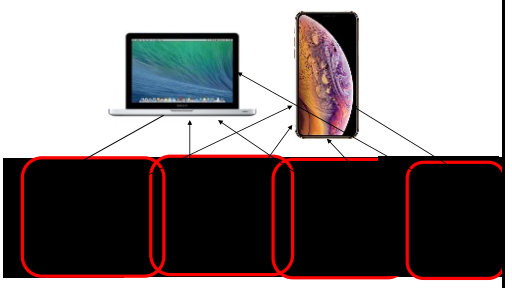
Different participants (Between-Subject Design)- single group of participants is allocated randomly to the experimental conditions.

Same participants (Within-Subject Design) - all participants appear in both conditions.

Matched participants (Pair-Wise Design) - participants are matched in pairs, e.g., based on expertise, gender, etc.

Within-Subject Experiment

Between-Subject Experiment



Matched Pair Wise Experiment

Different, Same, Matched - Design

Design	Advantages	Disadvantages
Different (Between-Subject)	No order effects	Many subjects & individual differences a problem
Same (Within-Subject)	Few individuals, no individual differences	Counter-balancing needed because of ordering effects
Matched (Pair-Wise)	Same as different participants but individual differences reduced	Cannot be sure of perfect matching on all differences

Usability Testing vs. Experiments

Usability testing

- Improve products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments for research

- Discover knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions
- Experimental design
- Scientific report to scientific community

Field Studies

- Field studies are done in natural settings.
- "In the wild" is a term for prototypes being used freely in natural settings.
- Aim to understand what users do naturally and how technology impacts them.
- Field studies are used in product design to:
  - identify opportunities for new technology;
  - determine design requirements;
  - decide how best to introduce new technology;
  - evaluate technology in use.

An in the wild study: UbiFit Garden

Figure 14.8 UbiFit Garden's glanceable display: (a) at the beginning of the week (small butterflies indicate recent goal attainments; the absence of flowers means no activity this week); (b) a garden with workout variety; (c) the display on a mobile phone (the large butterfly indicates this week's goal was met)

Source: From Consolvo, S., McDonald, D.W., Toscos, T. *et al* (2008) "Activity sensing in the wild: a field trial of UbiFit garden". In: *Proceedings of CHI 2008*, ACM Press, New York, p. 1799.

A/B Testing

A type of large-scale experiment

Offers another way to evaluate a website, application or app running on a mobile device

Often used for evaluating changes in design of social media applications

Compares how two groups of users perform on two versions of a design

A/B Testing *(continued)*

Can involve thousands of people

May create ethical dilemmas if users don't know they are part of the test

Care is needed to ensure that other issues are not affecting users' behavior

A/B Testing *Example*

Imagine you manage an e-commerce website and want to increase the number of users who add items to their shopping cart. You hypothesize that a more prominent "Add to Cart" button will improve this metric.

Control (Version A): The current "Add to Cart" button is small and blue.

Variant (Version B): The new "Add to Cart" button is larger and red.

You randomly assign half of your website visitors to see the blue button and the other half to see the red button. After running the test for a week, you analyze the data and find that the red button has a significantly higher click-through rate. Based on this result, you decide to implement the red button for all users.

Predictive models

Provide a way of evaluating products or designs without directly involving users

Less expensive than testing with users and some kinds of inspections

Usefulness limited to systems with predictable tasks, for example, voicemail systems, smartphones, and dedicated mobile devices

Based on knowledge of expert error-free behavior

Fitts' Law (1954)

Fitts' Law predicts that the time to point at an object using a device is a function of the distance from the target object and the object's size

The further away and the smaller the object, the longer the time to locate it and point to it

It is particularly useful for determining where on a screen to position an object

Fitts' Law is useful for evaluating systems for which the time to locate an object is important, for example, smartphones, handhelds, and mobile devices

Fitts' Law (1954)

$$T = a + b \log_2 \left(\frac{2D}{W} + 1\right)$$

where:

- **T** is the average time to complete the movement.
- **a** and **b** are empirical constants that depend on the specific characteristics of the device and the user.
- **D** is the distance from the starting point to the center of the target.
- **W** is the width of the target along the axis of movement.

Key points

Usability testing takes place in controlled usability labs or temporary labs, focuses on performance measures, eg. how long and how many errors are made when completing a set of predefined tasks. Indirect observation (video and keystroke logging), user satisfaction questionnaires and interviews are also collected.

Affordable, remote testing systems are more portable than usability labs. Many also contain mobile eye-tracking and other devices.

Experiments test a hypothesis by manipulating certain variables while keeping others constant.

The experimenter controls independent variable(s) in order to measure dependent variable(s).

Field studies are evaluation studies that are carried out in natural settings to discover how people interact with technology in the real world.

Field studies that involve the deployment of prototypes or technologies in natural settings may also be referred to as 'in the wild'.

That's it for today!