EECS 4441 Human-Computer Interaction

Topic #4: Empirical Research Methods for HCI

I. Scott MacKenzie

York University, Canada

LASSONDE SCHOOL OF ENGINEERING | YORK UNIVERSITÉ UNIVERSITY | EECS ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

What is Empirical Research?

- Empirical Research is…
  - investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws
  - based on observation or experience; capable of being verified or disproved by observation or experiment
- In HCI, we focus on "relevant to phenomena surrounding humans interacting with computers"

see http://www.merriam-webster.com/dictionary

Why do Empirical Research?

- We conduct empirical research to…
  - Answer (and raise!) questions about new or existing user interface designs or interaction techniques
  - Find cause-and-effect relationships
  - Transform baseless opinions into informed opinions supported by evidence
  - Develop or test models that *describe* or *predict* behavior (of humans interacting with computers)

How do we do Empirical Research?

- We conduct empirical research through…
  - A program of inquiry conforming to the *scientific method*
- The scientific method is…
  - The principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses

Non-experimental Research

- Also important in HCI
- Tends to be qualitative, rather than quantitative
- Observation important (measurement less so)
- Motivation
  - Reasons underlying human behaviour
  - Why, as opposed to what or how
- Focus
  - Human thought, emotion, sensation, reflection, expression, sentiment, opinion, outlook, manner, approach, strategy, etc.
- How
  - Interviews, case studies, field studies, focus groups, think aloud protocols, story telling, walkthroughs, cultural probes, etc.

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Observations and Measurements

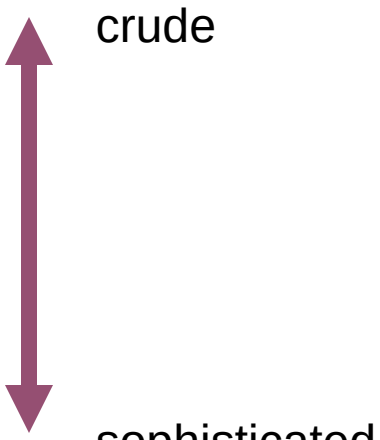
- Observations are gathered…
  - Manually (human observers)
  - Automatically (computers, software, cameras, sensors, etc.)
- A measurement is a recorded observation

When you cannot measure, your knowledge is of a meager and unsatisfactory kind.

Kelvin, 1883

Scales of Measurement

- Nominal
- Ordinal
- Interval
- Ratio



Nominal Data

- Nominal data (aka categorical data) are arbitrary codes assigned to attributes; e.g.,
  - M = male, F = female
  - 1 = mouse, 2 = touchpad, 3 = pointing stick
- Obviously, the statistical mean cannot be computed on nominal data
- Usually it is the count that is important
  - “Are females or males more likely to…”
  - “Do left or right handers have more difficulty with…”
  - Note: The count itself is a ratio-scale measurement

Nominal Data Example In HCI

- Observe students “on the move” on university campus
- Code and count students by…
  - Gender (male, female)
  - Mobile phone usage (not using, using)

Gender	Mobile Phone Usage		Total	%
	Not Using	Using		
Male	683	98	781	51.1%
Female	644	102	746	48.9%
Total	1327	200	1527	
%	86.9%	13.1%		

Real Data !

Ordinal Data

- Ordinal data associate order or rank to an attribute
- The attribute is any characteristic or circumstance of interest; e.g.,
  - Users try three different GPS systems for a period of time, then rank them: 1st, 2nd, 3rd choice
- More sophisticated than nominal data
  - Comparisons of “greater than” or “less than” possible

Ordinal Data Example in HCI

How many email messages do you receive each day?

1. None (I don't use email)
2. 1-5 per day
3. 6-25 per day
4. 26-100 per day
5. More than 100 per day

Interval Data

- Equal distances between adjacent values
- But, no absolute zero
- Classic example: temperature (°F, °C)
- Statistical mean possible
  - E.g., the mean midday temperature during July
- Ratios not possible
  - Cannot say 10 °C is twice 5 °C

Interval Data Example in HCI

- Questionnaires often solicit a level of agreement to a statement
- Responses on a Likert scale
- Likert scale characteristics:
  1. Statement soliciting level of agreement
  2. Responses are symmetric about a neutral middle value
  3. Gradations between responses are equal (more-or-less)
- Assuming “equal gradations”, the statistical mean is valid (and related statistical tests are possible)

Interval Data Example in HCI (2)

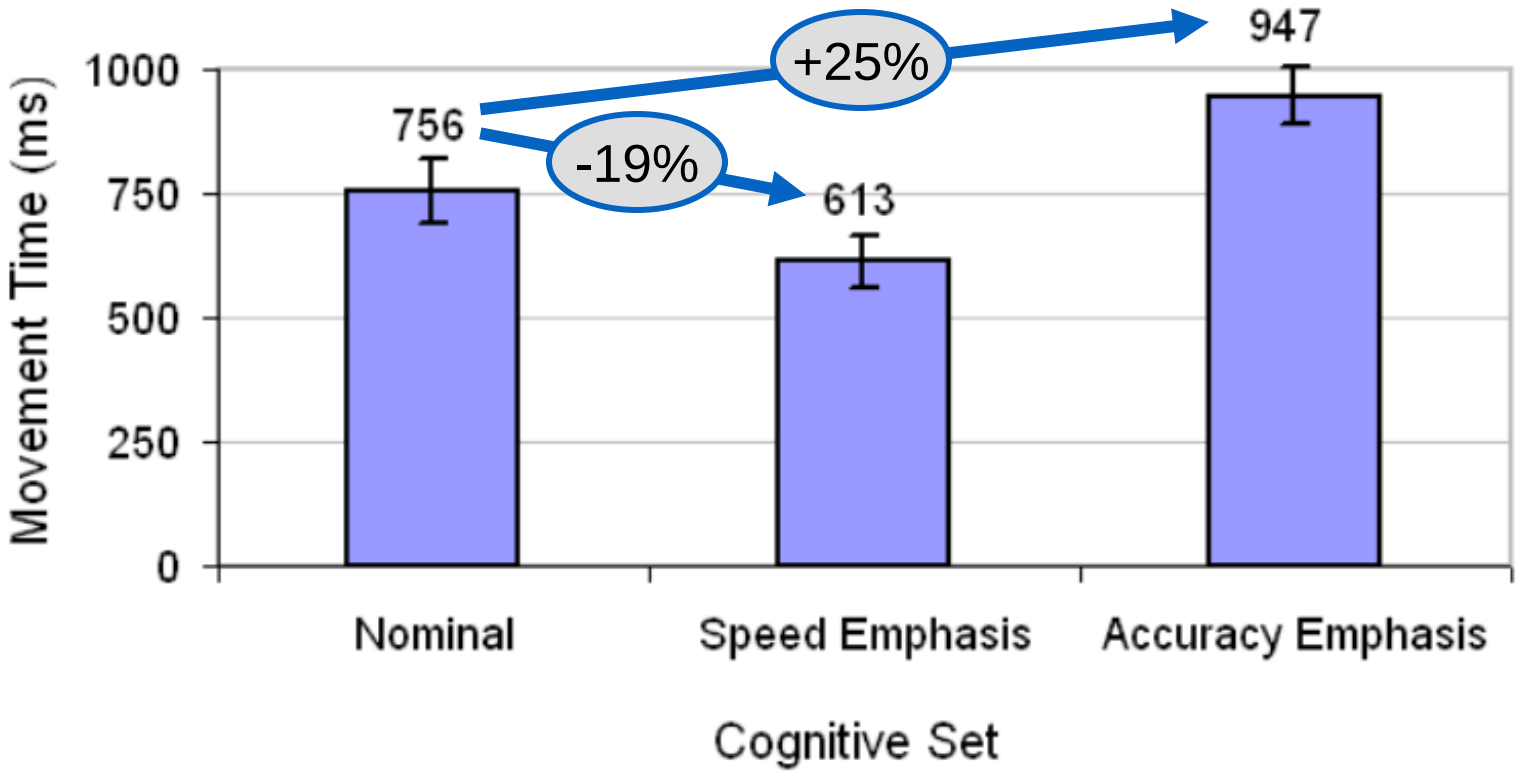
Please indicate your level of agreement with the following statements.

	Strongly disagree	Mildly disagree	Neutral	Mildly agree	Strongly agree
It is safe to talk on a mobile phone while driving.	1	2	3	4	5
It is safe to compose a text message on a mobile phone while driving.	1	2	3	4	5
It is safe to read a text message on a mobile phone while driving.	1	2	3	4	5

Ratio Data

- Most sophisticated of the four scales of measurement
- Preferred scale of measurement
- Absolute zero, therefore many calculations possible
- Summaries and comparisons are strengthened
- A “count” is a ratio-scale measurement
  - E.g., “time” (the number of seconds to complete a task)
- Enhance counts by adding further ratios where possible
  - Facilitates comparisons
  - Example – a 10-word phrase was entered in 30 seconds
    - Bad: *t* = 30 seconds (0.5 minutes)
    - Good: Entry rate = 10 / 0.5 = 20 wpm (words-per-minute)

Ratio Data Example in HCI



$$F_{2,34} = 372.7,\ p < .0001$$

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Research Questions

- We conduct empirical research to answer (and raise!) questions about UI designs or interaction techniques
- Consider the following questions:
  - Is it viable?
  - Is it better than current practice?
  - Which design alternative is best?
  - What are the performance limits?
  - What are the weaknesses?
  - Does it work well for novices?
  - How much practice is required?

Testable Research Questions

- Preceding questions, while unquestionably relevant, are not testable
- Try to re-cast as testable questions (even though the new question may appear less important)
- Scenario…
  - You have invented a new optimized keyboard (NOK) for smart phones, and you think it's pretty good. In fact, you think it is better than the Qwerty soft keyboard (QSK). You decide to undertake a program of empirical enquiry to evaluate your invention. What are your research questions?

Research Questions (2)

- Very weak

  Is the NOK any good?

- Weak

  Is the NOK better than QSK?

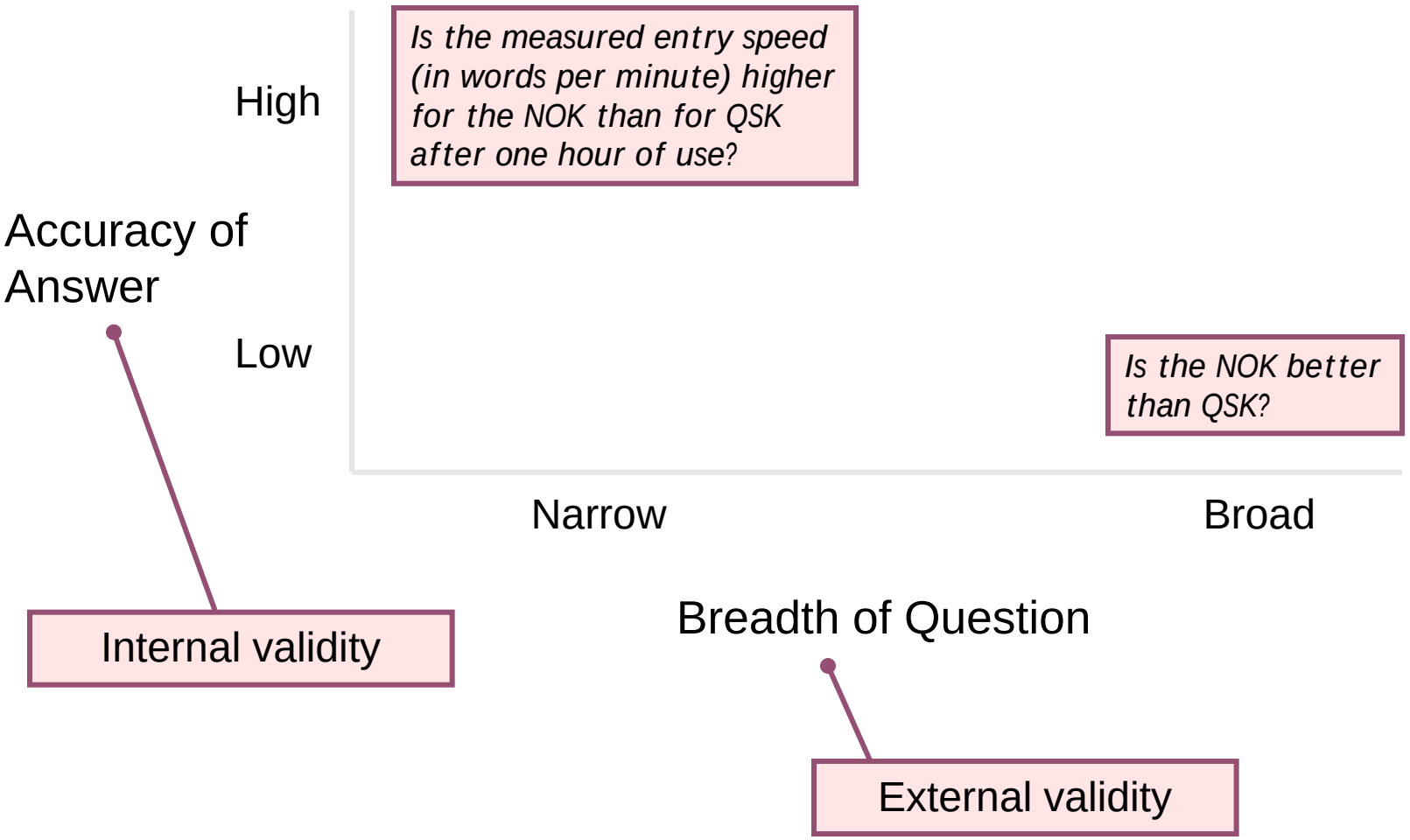
- Better

  Is the NOK faster than QSK?

- Better still

  Is the measured entry speed (in words per minute) higher for the NOK than for QSK after one hour of use?

A Tradeoff


Is the measured entry speed (in words per minute) higher for the NOK than for QSK after one hour of use?
High
Accuracy of Answer
Low
Is the NOK better than QSK?
Narrow
Broad
Breadth of Question
Internal validity
External validity

Internal Validity

- Definition:
  - The extent to which the effects observed are due to the test conditions (e.g., NOK vs. QSK)
- Statistically, this means…
  - Differences (in the means) are due to inherent properties of the test conditions
  - Variances are due to participant differences ("pre-dispositions")
  - Other potential sources of variance are controlled or exist equally or randomly across the test conditions

External Validity

- Definition:
  - The extent to which results are generalizable to other ***people*** and other ***situations***
- People
  - The participants are ***representative*** of the broader intended population of users
- Situations
  - The ***test environment*** and ***experimental procedures*** are representative of real world situations where the interface or technique will be used

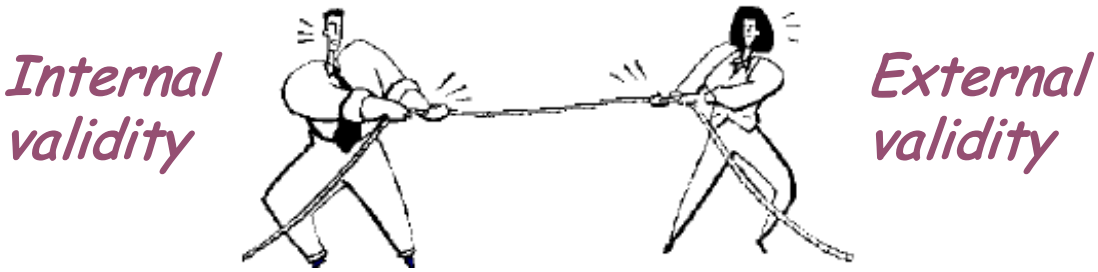
Test Environment Example

- Scenario…
  - You wish to compare two input devices for remote pointing (e.g., at a projection screen)
- External validity is improved if the test environment mimics expected usage
- Test environment should probably…
  - Use a large display or projection screen (not a desktop monitor)
  - Position participants at a significant distance from screen (rather than close up)
  - Have participants stand (rather than sit)
  - Include an audience!
- But… is internal validity compromised?

Experimental Procedure Example

- Scenario…
  - You wish to compare two text entry techniques for mobile devices
- External validity is improved if the experimental procedure mimics expected usage
- Test procedure should probably have participants…
  - Enter representative samples of text (e.g., phrases containing letters, numbers, punctuation, etc.)
  - Edit and correct mistakes as they normally would
- But… is internal validity compromised?

The Tradeoff



- There is tension between internal and external validity
- The more the test environment and experimental procedures are “relaxed” (to mimic real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as pondering, distractions, or secondary tasks

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Experiment Terminology (Part 1)

- Terms to know
  - Participant
  - Independent variable (test conditions)
  - Dependent variable (measured behaviors)
  - Control variable
  - Random variable
  - Confounding variable
  - Within subjects vs. between subjects
  - Counterbalancing
  - Latin square

Participant

- The people participating in an experiment are referred to as ***participants*** (the term ***subjects*** is not commonly used)
- When referring specifically to the experiment, use ***participants*** (e.g., “all ***participants*** exhibited a high error rate…”)
- General discussion on the problem or conclusions may use other terms (e.g., “these results suggest that ***users*** are less likely to…”)
- Report the selection criteria and give relevant demographic information or prior experience

Independent Variable

- An ***independent variable*** is a circumstance that is manipulated through the design of the experiment
- It is “independent” because it is independent of participant behavior (i.e., there is nothing a participant can do to influence an independent variable)
- Examples include interface, device, feedback mode, button layout, visual layout, gender, age, expertise, etc.
- The terms ***independent variable*** and ***factor*** are synonymous

Test Conditions

- The levels, values, or settings for an independent variable are the ***test conditions***
- Provide a name for both the ***factor*** (***independent variable***) and its ***levels*** (***test conditions***)
- Examples

Factor	***Test Conditions (Levels)***
Device	mouse, trackball, joystick
Feedback mode	audio, tactile, none
Task	pointing, dragging
Visualization	2D, 3D, animated
Search interface	Google, custom

Dependent Variable

- A ***dependent variable*** is any measurable aspect of the interaction involving an independent variable
- Examples include task completion time, speed, accuracy, error rate, throughput, target re-entries, task retries, presses of backspace, etc.
- Give a name to the dependent variable, separate from its units (e.g., “Text Entry Speed” is a dependent variable with units “words per minute”)
- Make sure you clearly define all dependent variables
- Research must be reproducible!

Topics

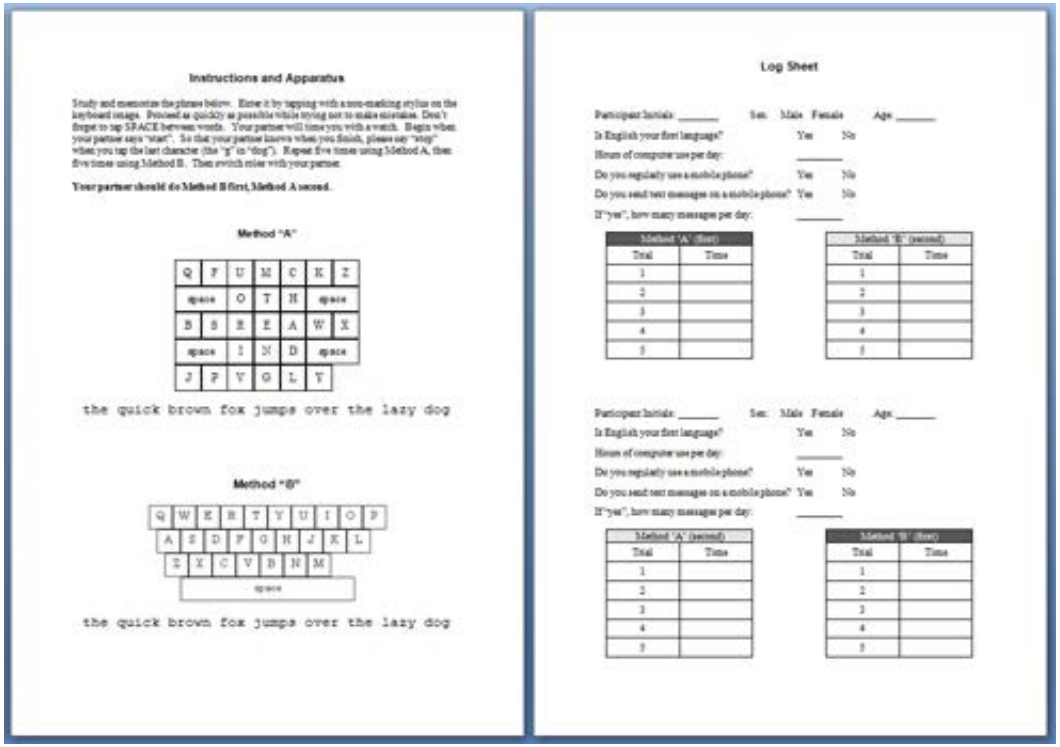
- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Group Participation [1]

- At this point in the course, attendees are divided into groups of two to participate in a real user study
- A three-page handout is distributed to each group (see next slide)
- Read the instructions on the first page and discuss the procedure with your partner
- Your instructor will provide additional information

[1]This section may be shortened depending on the time available

Handout (2 pages)



Instructions and Apparatus

Study and memorize the phrase below. Enter it by tapping with a non-marking stylus on the keyboard image. Proceed as quickly as possible while trying not to make mistakes. Don't forget to tap SPACE between words. Your partner will time you with a watch. Begin when your partner says "start". So that your partner knows when you finish, please say "stop" when you tap the last character (the "g" in "dog"). Repeat five times using Method A, then five times using Method B. Then switch roles with your partner.

Your partner should do Method B first, Method A second.

Method "A"

Q	F	U	M	C	K	Z
space	O	T	H	space		
B	S	R	E	A	W	X
space	I	N	D	space		
J	P	V	G	L	Y	

the quick brown fox jumps over the lazy dog

Method "B"

Q	W	E	R	T	Y	U	I	O	P
A	S	D	F	G	H	J	K	L	
Z	X	C	V	B	N	M			
space									

the quick brown fox jumps over the lazy dog

Log Sheet

Participant Initials: _____ Sex: Male Female Age: _____

Is English your first language? Yes No

Hours of computer use per day: _____

Do you regularly use a mobile phone? Yes No

Do you send text messages on a mobile phone? Yes No

If "yes", how many messages per day: _____

Method 'A' (first)	
Trial	Time
1	
2	
3	
4	
5	

Method 'B' (second)	
Trial	Time
1	
2	
3	
4	
5	

Participant Initials: _____ Sex: Male Female Age: _____

Is English your first language? Yes No

Hours of computer use per day: _____

Do you regularly use a mobile phone? Yes No

Do you send text messages on a mobile phone? Yes No

If "yes", how many messages per day: _____

Method 'A' (second)	
Trial	Time
1	
2	
3	
4	
5	

Method 'B' (first)	
Trial	Time
1	
2	
3	
4	
5	

Full-size copies of the handout pages will be distributed during the course. The pages are also contained in an appendix to this package.

Do the Experiment

- The experiment is performed
- This takes about 30 minutes
- After the experiment… break time
- The instructor and an assistant will transcribe the tabulated data into a ready-made spreadsheet
- Results are instantaneous
- After the break… (next slide)

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Experiment Design

- ***Experiment design*** is the process of deciding what variables to use, what tasks and procedures to use, how many participants to use and how to solicit them, and so on
- Let's continue with some terminology…

Experiment Terminology (Part 2)

- Terms to know
  - Participant
  - Independent variable (test conditions)
  - Dependent variable (measured behaviors)
  - Control variable
  - Random variable
  - Confounding variable
  - Within subjects vs. between subjects
  - Counterbalancing
  - Latin square

Control Variable

- Circumstance (not under investigation) that is kept constant to test the effect of an independent variable
- More control means the experiment is less generalizable (i.e., less applicable to other people and other situations)
- Consider an experiment on the effect of font color and background color on reader comprehension
  - Independent variables: font color, background color
  - Dependent variables: comprehension test scores
  - Control variables
    - Font size (e.g., 12 point)
    - Font family (e.g., Times)
    - Ambient lighting (e.g., fluorescent, fixed intensity)
    - Etc.

Random Variable

- Circumstance that is allowed to vary randomly
- More variability is introduced in the measures (that's bad!), but the results are more generalizable (that's good!)
- Consider an experiment comparing whether a user's stance affects performance while playing *Guitar Hero*
  - Independent variable: stance (standing, sitting)
  - Dependent variable: score on songs
  - Random variables
    - Prior experience playing a real musical instrument
    - Prior experience playing *Guitar Hero*
    - Amount of coffee consumed prior to testing
    - Etc.

Confounding Variable

- Circumstance that varies systematically with an independent variable
- Should be controlled or randomized to avoid misleading results
- Consider a study comparing the target selection performance of a mouse and a gamepad where all participants are mouse experts, but gamepad novices
  - Mouse performance will likely be higher, but…
  - “Prior experience” is a confounding variable
  - No reliable conclusions can be made

How Many Participants

- Short answer
  - Use the same number of participants as used in similar research
- Too many participants…
  - and you get statistically significant results for differences of no *practical* significance
- Too few participants…
  - and you fail to get statistically significant results when there really is an inherent difference between the test conditions

Within Subjects, Between Subjects

- The administering of levels of a factor is either ***within subjects*** or ***between subjects***
- If each participant is tested on each level, the factor is ***within subjects***
- If each participant is tested on only one level, the factor is ***between subjects***. (In this case, a separate group of participants is used for each level.)
- The terms ***repeated measures*** and ***within subjects*** are synonymous.

Within vs. Between Subjects

- Question: Should a factor be assigned within subjects or between subjects?
- Answer: It depends!
- Sometimes a factor must be between subjects (e.g., gender, age)
- Sometimes a factor must be within subjects (e.g., session, block)
- Sometimes there is a choice
- Within subjects advantage – the variance due to participants' pre-dispositions should be the same across test conditions (cf. between subjects)
- Between subjects advantage – avoids interference effects (e.g., typing on two different layouts of keyboards)

Counterbalancing

- For within-subjects designs, participants' performance may improve with practice as they progress from one test condition to the next. Thus, participants may perform better on the second condition simply because they benefited from practice on the first. We don't want this.
- To compensate, the order of presenting conditions is ***counterbalanced***
- Participants are divided into ***groups***, and a different order of administration is used for each group
- The order is best governed by a ***Latin Square*** (next slide)
- ***Group***, then, is a between subjects factor (Was there an effect for group? Hopefully not!)

Latin Square

- The defining characteristic of a Latin Square is that each condition occurs only once in each row and column
- Examples:

3 X 3 Latin Square

A	B	C
B	C	A
C	A	B

4 x 4 Latin Square

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

4 x 4 Balanced Latin Square

A	B	C	D
B	D	A	C
D	C	B	A
C	A	D	B

Note: In a ***balanced Latin Square*** each condition both precedes and follows each other condition an equal number of times

Succinct Statement of Design

- "***3 x 2 repeated-measures design***" refers to an experiment with two factors, having ***three levels*** on the first, and ***two levels*** on the second. There are ***six test conditions*** in total. Both factors are repeated measures, meaning all participants were tested on all conditions
- Note: A mixed design is also possible
  - In a mixed design, the levels for one factor are administered to all participants (within subjects) while the levels for another factor are administered to separate groups of participants (between subjects).

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper

Answering Research Questions

- We want to know if the measured performance on a variable (e.g., entry speed) is different between test conditions, so…
  - We conduct a user study and measure the performance on each test condition with a group of participants
  - For each test condition, we compute the mean score over the group of participants
  - Then what?

Answering Research Questions (2)

- Four questions:
  1. Is there a difference?
  2. Is the difference large or small?
  3. Is the difference statistically significant (or is it due to chance)?
  4. Is the difference of practical significance?
- Q1 – obvious (some difference is likely)
- Q2 – statistics can't help (Is a 5% difference large or small?)
- Q3 – statistics can help
- Q4 – statistics can't help (Is a 5% difference useful? People resist change!)
- The basic statistical tool for Q3 is the analysis of variance (ANOVA)

Null Hypothesis

- Formally speaking, a research question is not a question. It is a statement called the *null hypothesis.*
- Example:

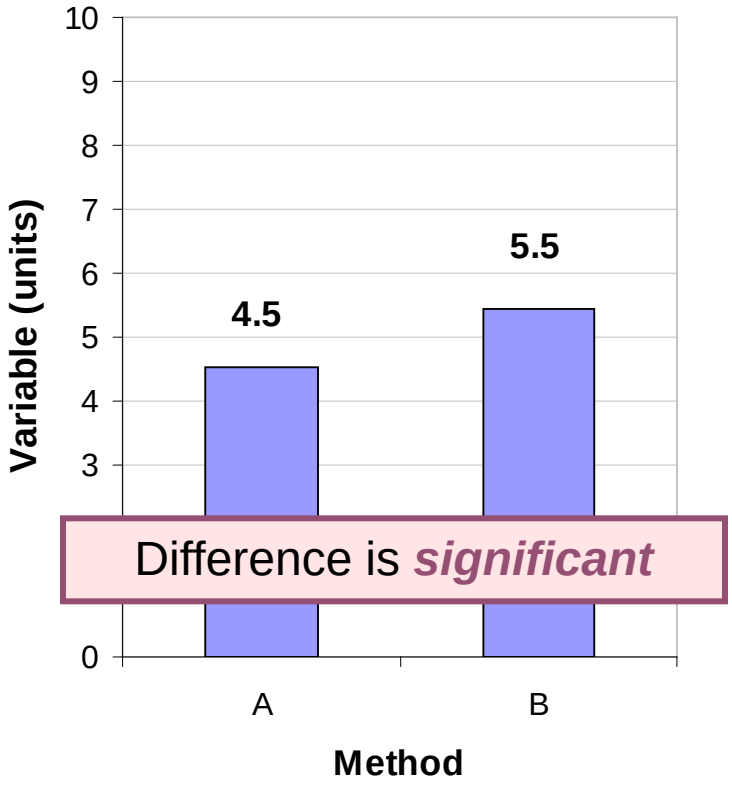
> There is no difference in entry speed between Method A and Method B.

- Assumption of “no difference”
- Research seeks to reject the null hypothesis
- Please bear in mind, with experimental research…
  - We gather evidence
  - We do not prove things

Analysis of Variance

- It is interesting that the test is called an analysis of ***variance***, yet it is used to determine if there is a significant difference between the ***means***.
- How is this?

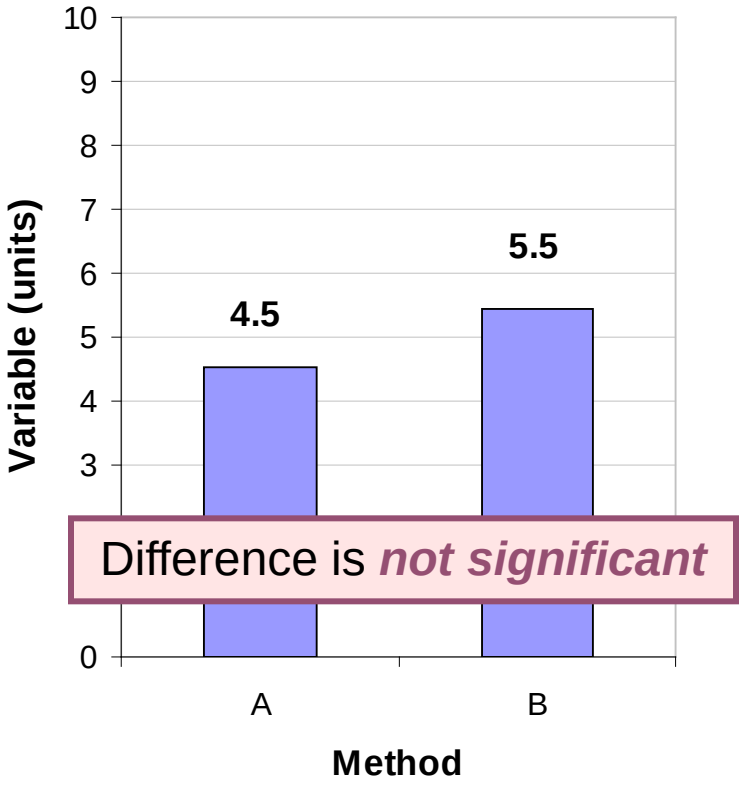
Example #1



Difference is ***significant***

"Significant" implies that in all likelihood the difference observed is due to the test conditions (Method A vs. Method B).

Example #2

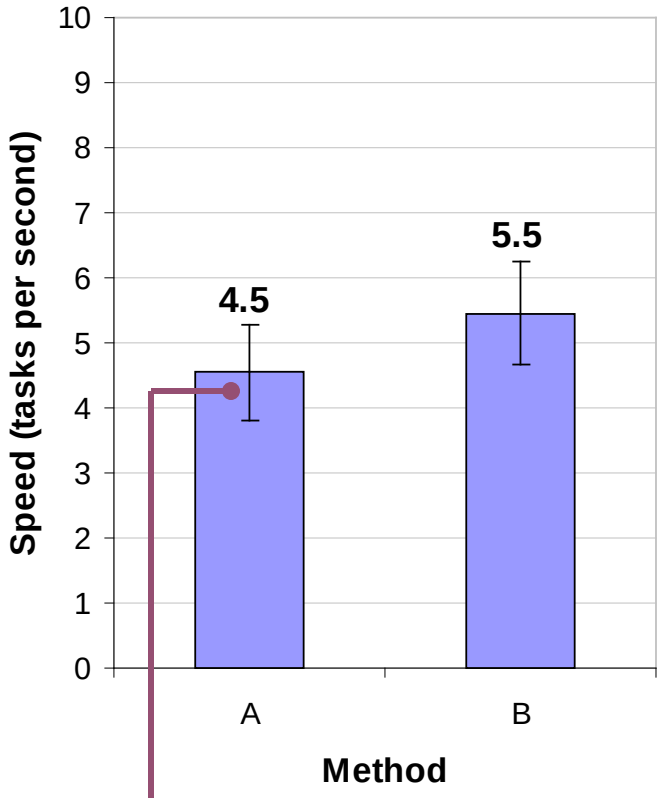


Difference is ***not significant***

"Not significant" implies that the difference observed is likely due to chance.

File: AnovaDemo.xls

Example #1 - Details



Example #1		
Participant	Method	
	A	B
1	5.3	5.7
2	3.6	4.6
3	5.2	5.1
4	3.3	4.5
5	4.6	6.0
6	4.1	7.0
7	4.0	6.0
8	5.0	4.6
9	5.2	5.5
10	5.1	5.6
Mean	4.5	5.5
SD	0.73	0.78

Error bars show
±1 standard deviation

Note: *SD* is the square root of the variance

Example #1 - ANOVA

ANOVA Table for Speed

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	5.839	.649				
Method	1	4.161	4.161	8.443	.0174	8.443	.741
Method * Subject	9	4.435	.493				

Probability of obtaining the observed data if the null hypothesis is true

Reported as…

$F_{1,9} = 8.443, p < .05$

Thresholds for "p"

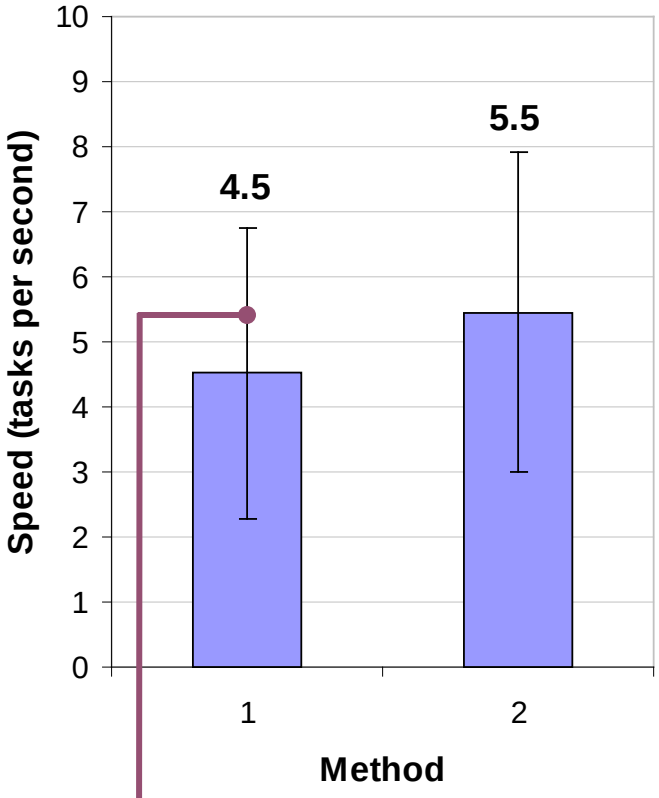
- .05
- .01
- .005
- .001
- .0005
- .0001

How to Report an *F*-statistic

There was a significant effect of input method on entry speed ($F_{1,9} = 8.44$, $p < .05$).

- Notice in the parentheses
  - Uppercase for *F*
  - Lowercase for *p*
  - Italics for *F* and *p*
  - Space both sides of equal sign
  - Space after comma
  - Space on both sides of less-than sign
  - Degrees of freedom are subscript, plain, smaller font
  - Three significant figures for *F* statistic
  - No zero before the decimal point in the *p* statistic (except in Europe)

Example #2 - Details



Error bars show
±1 standard deviation

Example #2		
Participant	Method	
	A	B
1	2.4	6.9
2	2.7	7.2
3	3.4	2.6
4	6.1	1.8
5	6.4	7.8
6	5.4	9.2
7	7.9	4.4
8	1.2	6.6
9	3.0	4.8
10	6.6	3.1
Mean	4.5	5.5
SD	2.23	2.45

Example #2 – ANOVA

ANOVA Table for Speed

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	37.017	4.113				
Method	1	4.376	4.376	.634	.4462	.634	.107
Method * Subject	9	62.079	6.898				

Probability of obtaining the observed data if the null hypothesis is true

Reported as…

$F_{1,9} = 0.634$, ns

Note: For non-significant effects, use "ns" if $F < 1.0$, or "$p > .05$" if $F > 1.0$.

Main Effects vs. Interaction Effects

- If there are two independent variables, check for *main effects* (2) and an *interaction effect*
- E.g., Effect of "Feedback Mode" (4 levels) and "Block" (4 blocks) on error rate (%)

ANOVA Table for Error Rate (%)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	12	44.105	3.675				
Feedback	3	15.895	5.298	4.918	.0058	14.753	.885
Feedback * Subject	36	38.786	1.077				
Block	3	17.541	5.847	4.980	.0054	14.939	.890
Block * Subject	36	42.271	1.174				
Feedback * Block	9	16.144	1.794	1.727	.0914	15.542	.758
Feedback * Block * Subject	108	112.185	1.039				

- http://www.yorku.ca/mack/chi03d.html (or check Anova2 API)

Reporting an F-statistic – Revisited

- Default format mentions both the independent variable and the dependent variable:

> "The effect of *independent_variable* on *dependent_variable* was statistically significant (F-STATISTIC)."

- Example on next slide

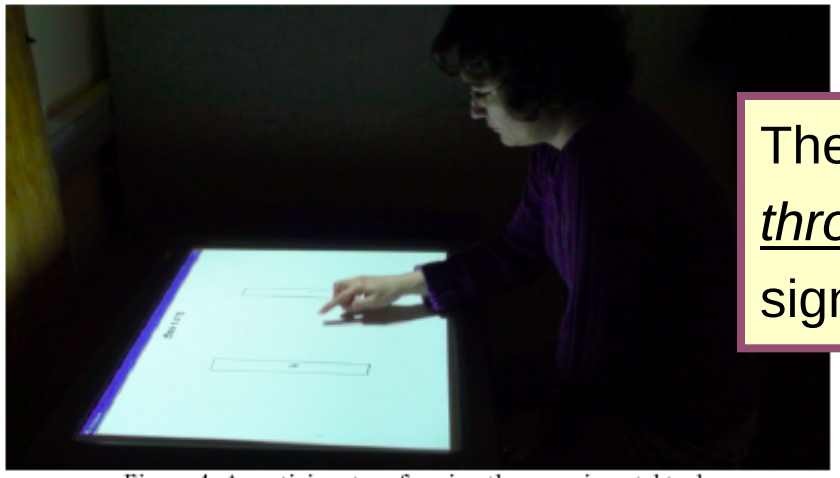


Figure 4. A participant performing the experimental task

RESULTS AND DISCUSSION

Throughput

Touch interaction yielded a higher throughput compared to the mouse. The overall mean throughput for touch interaction was 5.52 bps, which was 44.1% higher than the 3.83 bps observed for the mouse. The effect of input technique on throughput was statistically significant ($F_{1,11}$ = 35.51, p < .0001). Although not as high as the throughput reported by Forlines et al. (2007) for touch input (discussed earlier), our throughput values were computed using a direct

The effect of *input technique* on *throughput* was statistically significant ($F_{1,11}$ = 35.51, p < .0001).

Independent variable:
Input technique

Dependent variable:
Throughput

Sasangohar, F., MacKenzie, I. S., & Scott, S. D. (2009). Evaluation of mouse and touch input for a tabletop display using Fitts' reciprocal tapping task. *Proceedings of HFES 2009*, pp. 839-843. Santa Monica, CA: Human Factors and Ergonomics Society.

Post Hoc Comparisons

- A significant *F*-test means at least one mean is different from at least one other mean
- Does not reveal which pairs of means are different
- For this a post hoc comparison test is used (aka pair-wise comparisons)
- Example tests
  - Sheffé, Tukey HSD, Fisher LSD, Bonferroni-Dunn

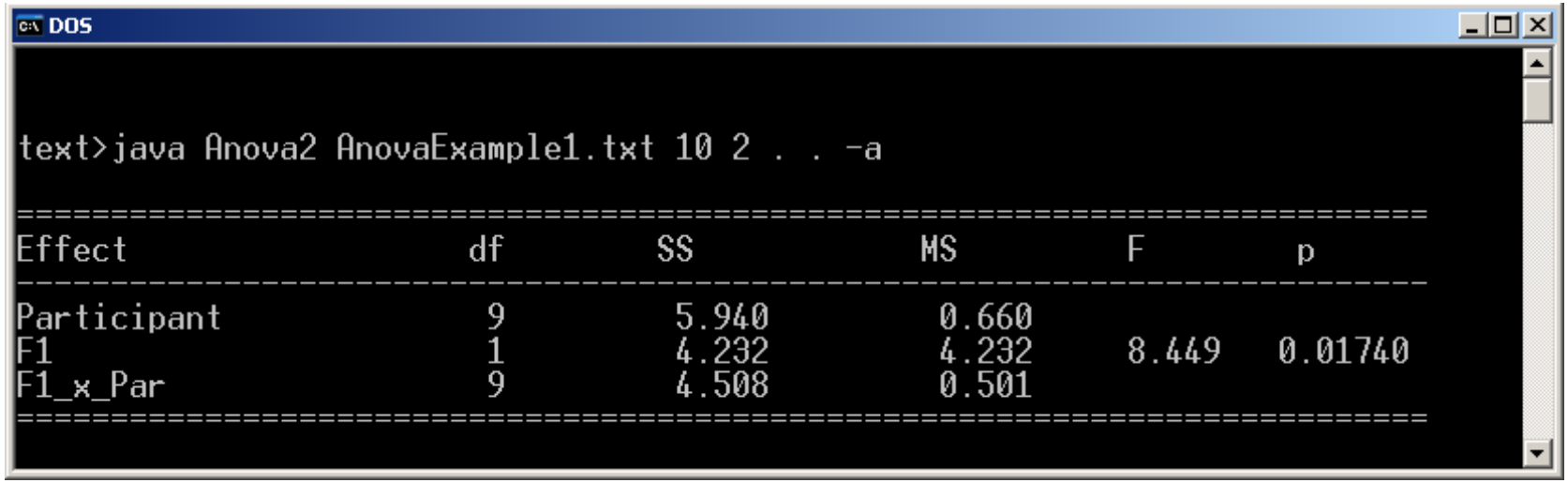
ANOVA Demos

- *StatView* (now sold as JMP, http://jmp.com)
  - Commercial statistics package
  - Input file: `AnovaExample1.svd`
- *Anova2*
  - Java program and its API are available (free download)
  - Input file: `AnovaExample1.txt`
- *PostHoc*
  - Java utility and its API are available (free download)

ANOVA Demos (2)

ANOVA Table for Speed

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	5.940	.660				
Method	1	4.232	4.232	8.449	.0174	8.449	.741
Method * Subject	9	4.508	.501				

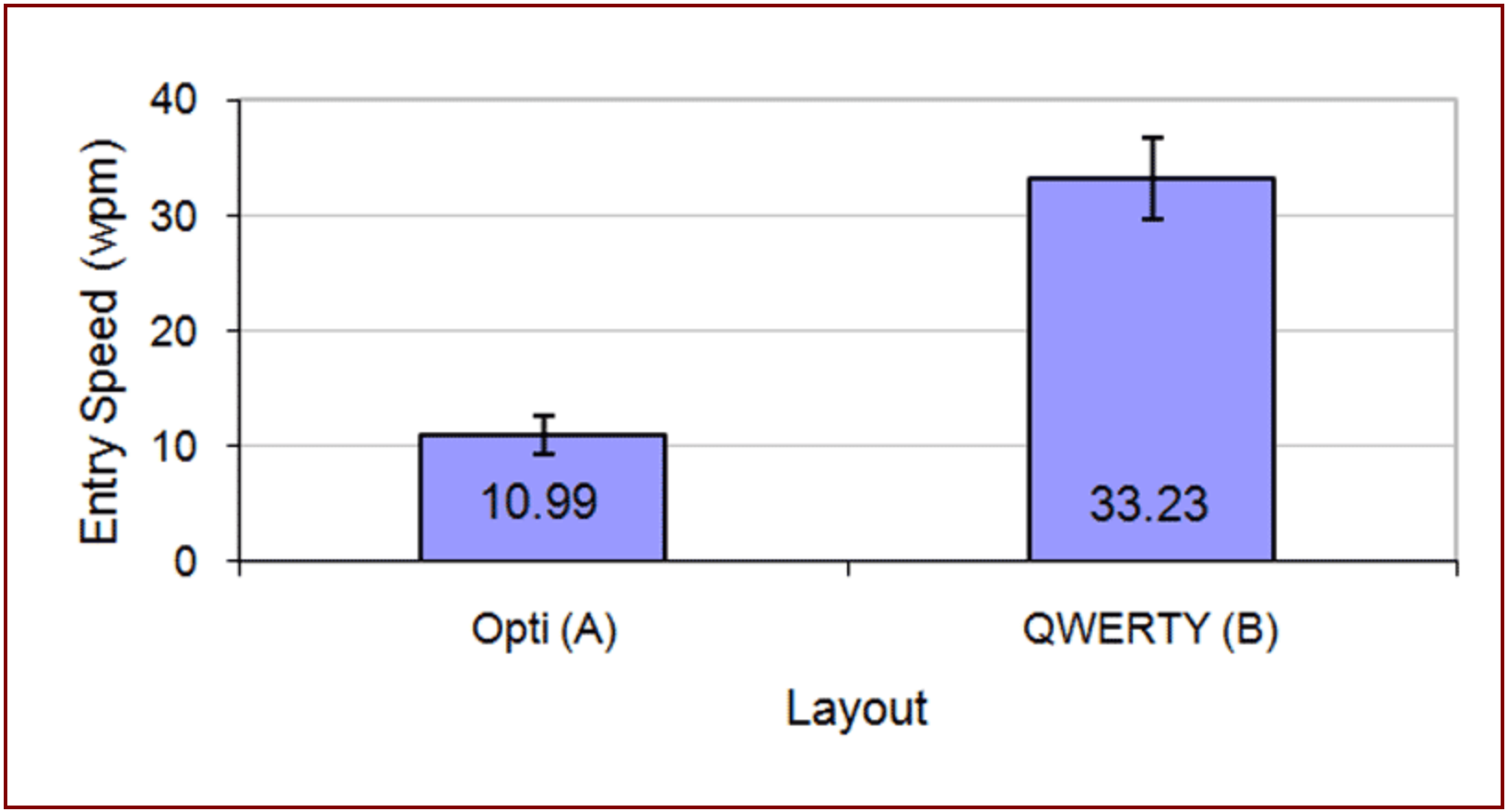


Group Participation Results

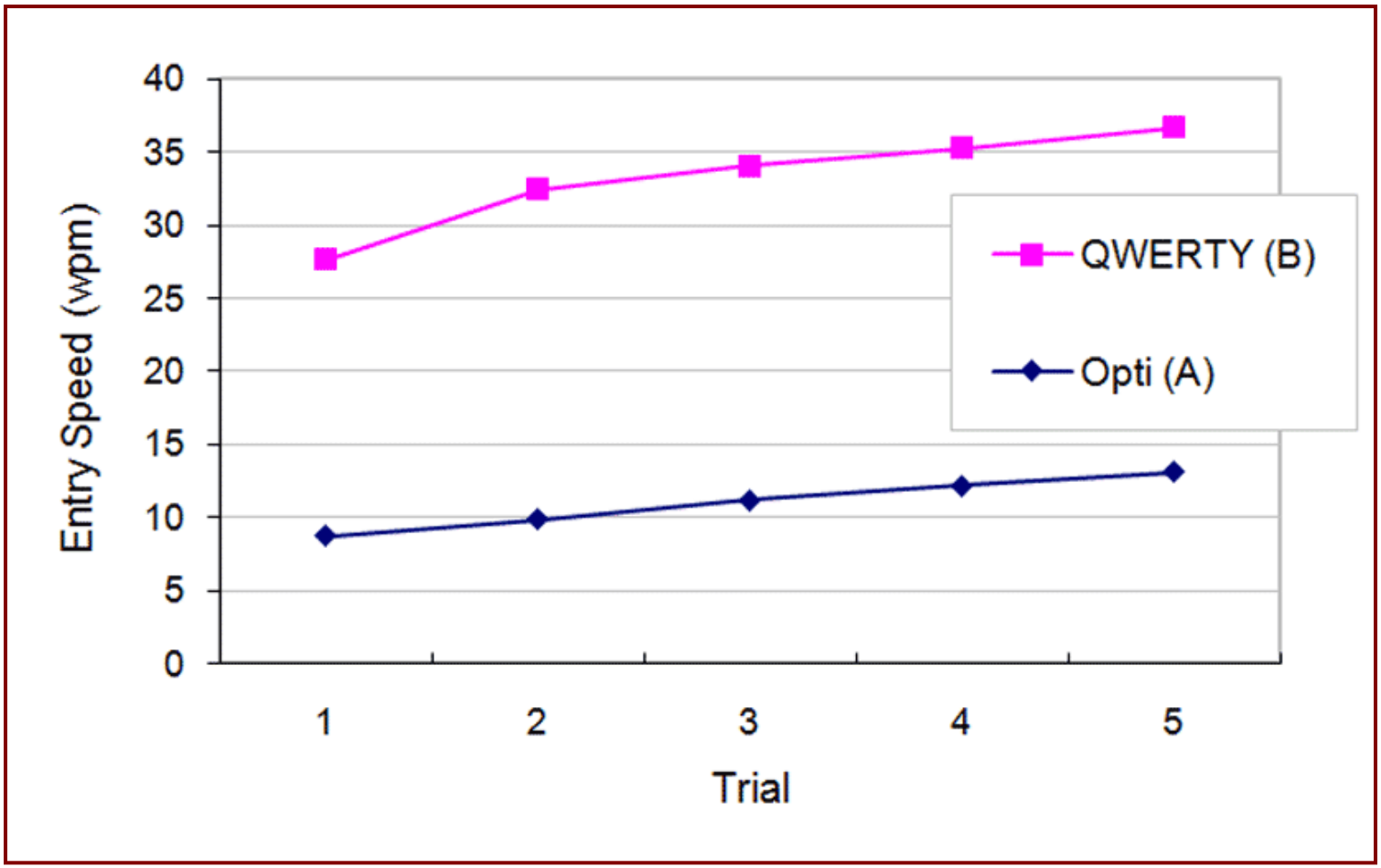
- The results presented in class are for the experiment conducted before the break
- The following results are from another run of the same experiment

Entry Time (seconds)												
Participant	Initials	Opti (A)					QWERTY (B)					Group
		1	2	3	4	5	1	2	3	4	5	
P1	al	92.0	94.0	84.0	68.0	93.0	23.0	19.0	17.0	17.0	15.0	1
P2	ig	65.0	63.0	55.0	49.0	41.0	18.0	15.0	14.0	14.0	13.0	1
P3	ma	54.0	44.0	38.0	38.0	32.0	19.0	17.0	17.0	15.0	19.0	1
P4	kw	65.0	71.0	57.0	61.0	51.0	23.0	19.0	19.0	19.0	18.0	1
P5	ja	40.0	33.0	31.0	29.0	28.0	19.0	17.0	19.0	17.0	16.0	1
P6	ej	66.0	65.0	47.0	52.0	46.0	20.0	17.0	17.0	15.0	14.0	1
P7	ml	50.0	49.0	40.0	36.0	31.0	22.0	18.0	16.0	16.0	14.0	1
P8	pa	68.0	47.0	46.0	35.0	34.0	17.0	13.0	12.0	16.0	12.0	1
P9	ul	86.0	83.0	56.0	46.0	45.0	29.0	19.0	18.0	17.0	15.0	1
P10	em	72.0	67.0	51.0	45.0	49.0	18.0	15.0	13.0	12.0	14.0	1
P11	pl	49.0	48.0	53.0	39.0	39.0	19.0	18.0	17.0	15.0	18.0	1
P12	bc	39.0	43.0	34.0	33.0	32.0	14.0	12.0	13.0	12.0	12.0	1
P13	as	54.0	44.0	41.0	38.0	41.0	17.0	14.0	12.0	13.0	13.0	2
P14	jj	75.0	65.0	55.0	71.0	53.0	21.0	17.0	17.0	19.0	16.0	2
P15	al	83.0	80.0	52.0	67.0	63.0	23.0	22.0	22.0	19.0	18.0	2
P16	sk	60.0	52.0	43.0	39.0	36.0	17.0	19.0	16.0	15.0	15.0	2
P17	jo	84.0	66.0	57.0	40.0	54.0	15.0	13.0	13.0	13.0	12.0	2
P18	hk	74.0	57.0	49.0	45.0	39.0	21.0	20.0	17.0	17.0	16.0	2
P19	mb	58.0	50.0	68.0	51.0	46.0	24.0	18.0	18.0	14.0	14.0	2
P20	jk	64.0	47.0	42.0	41.0	42.0	14.0	14.0	13.0	13.0	12.0	2
P21	ct	60.0	50.0	40.0	39.0	33.0	14.0	12.0	12.0	12.0	11.0	2
P22	hha	62.0	46.0	45.0	40.0	45.0	23.0	18.0	18.0	17.0	16.0	2
P23	ss	37.0	37.0	31.0	31.0	23.0	18.0	14.0	12.0	11.0	11.0	2
P24	ma	49.0	45.0	52.0	43.0	33.0	16.0	13.0	13.0	12.0	12.0	2

Entry Speed (wpm)												
Participant	Initials	Opti (A)					QWERTY (B)					Group
		1	2	3	4	5	1	2	3	4	5	
P1	al	5.61	5.49	6.14	7.59	5.55	22.43	27.16	30.35	30.35	34.40	1
P2	ig	7.94	8.19	9.38	10.53	12.59	28.67	34.40	36.86	36.86	39.69	1
P3	ma	9.56	11.73	13.58	13.58	16.13	27.16	30.35	30.35	34.40	27.16	1
P4	kw	7.94	7.27	9.05	8.46	10.12	22.43	27.16	27.16	27.16	28.67	1
P5	ja	12.90	15.64	16.65	17.79	18.43	27.16	30.35	27.16	30.35	32.25	1
P6	ej	7.82	7.94	10.98	9.92	11.22	25.80	30.35	30.35	34.40	36.86	1
P7	ml	10.32	10.53	12.90	14.33	16.65	23.45	28.67	32.25	32.25	36.86	1
P8	pa	7.59	10.98	11.22	14.74	15.18	30.35	39.69	43.00	32.25	43.00	1
P9	ul	6.00	6.22	9.21	11.22	11.47	17.79	27.16	28.67	30.35	34.40	1
P10	em	7.17	7.70	10.12	11.47	10.53	28.67	34.40	39.69	43.00	36.86	1
P11	pl	10.53	10.75	9.74	13.23	13.23	27.16	28.67	30.35	34.40	28.67	1
P12	bc	13.23	12.00	15.18	15.64	16.13	36.86	43.00	39.69	43.00	43.00	1
P13	as	9.56	11.73	12.59	13.58	12.59	30.35	36.86	43.00	39.69	39.69	2
P14	jj	6.88	7.94	9.38	7.27	9.74	24.57	30.35	30.35	27.16	32.25	2
P15	al	6.22	6.45	9.92	7.70	8.19	22.43	23.45	23.45	27.16	28.67	2
P16	sk	8.60	9.92	12.00	13.23	14.33	30.35	27.16	32.25	34.40	34.40	2
P17	jo	6.14	7.82	9.05	12.90	9.56	34.40	39.69	39.69	39.69	43.00	2
P18	hk	6.97	9.05	10.53	11.47	13.23	24.57	25.80	30.35	30.35	32.25	2
P19	mb	8.90	10.32	7.59	10.12	11.22	21.50	28.67	28.67	36.86	36.86	2
P20	jk	8.06	10.98	12.29	12.59	12.29	36.86	36.86	39.69	39.69	43.00	2
P21	ct	8.60	10.32	12.90	13.23	15.64	36.86	43.00	43.00	43.00	46.91	2
P22	hha	8.32	11.22	11.47	12.90	11.47	22.43	28.67	28.67	30.35	32.25	2
P23	ss	13.95	13.95	16.65	16.65	22.43	28.67	36.86	43.00	46.91	46.91	2
P24	ma	10.53	11.47	9.92	12.00	15.64	32.25	39.69	39.69	43.00	43.00	2
	Mean	8.72	9.82	11.18	12.17	13.06	27.63	32.43	34.07	35.29	36.71	
	SD	2.27	2.47	2.60	2.77	3.61	5.24	5.74	6.15	5.82	5.91	
					Min	5.49				*Min*	17.79	
					Max	22.43				*Max*	46.91	



Note: A *bar chart* is appropriate here because the data along the x-axis are categorical (i.e., nominal scale).



Note: A *line chart* is appropriate here because the data along the x-axis are continuous (i.e., ratio scale).

ANOVA Table for Entry Speed (wpm)

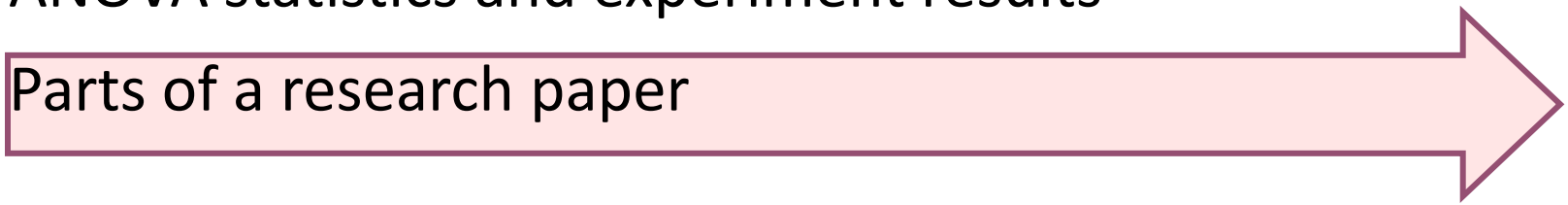
	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Group	1	73.737	73.737	.618	.4401	.618	.113
Subject(Group)	22	2624.205	119.282				
Layout	1	29664.381	29664.381	533.785	<.0001	533.785	1.000
Layout * Group	1	80.007	80.007	1.440	.2430	1.440	.199
Layout * Subject(Group)	22	1222.620	55.574				
Trial	4	1298.277	324.569	78.825	<.0001	315.300	1.000
Trial * Group	4	2.688	.672	.163	.9564	.653	.083
Trial * Subject(Group)	88	362.348	4.118				
Layout * Trial	4	172.752	43.188	10.706	<.0001	42.823	1.000
Layout * Trial * Group	4	10.887	2.722	.675	.6113	2.699	.207
Layout * Trial * Subject(Group)	88	354.997	4.034				

- Layout effect is significant ($F_{1,22}$ = 533.8, p < .0001)
- Trial effect is significant ($F_{4, 88}$ = 78.8, p < .0001)
- Layout by trial interaction effect is significant ($F_{4, 88}$ = 10.7, p < .0001)
- Group effect is not significant ($F_{1,22}$ = 0.62, *ns*)

Participant	Initials	Sex	Age	English as 1st language	Hours of computer use per day?	Do you regularly use a mobile phone?	Do you send text messages on a mobile phone?	If yes, how many messages per day?
P1	al	Male	43	No	10.0	Yes	Yes	8.0
P2	ig		35		7.0	Yes	n	0.0
P3	ma	female		Yes	8.0	Yes	Yes	5.0
P4	kw	female	33	No	8.0	Yes	Yes	2.5
P5	ja	Male	31	No	10.0	Yes	Yes	20.0
P6	ej	Male	42	Yes	10.0	Yes	Yes	20.0
P7	ml	female	41	No	8.0	Yes	Yes	5.0
P8	pa	Male	39	No	12.0	Yes	Yes	1.0
P9	ul	Male	36	No	10.0	Yes	Yes	3.0
P10	em	Male	45	Yes	8.0	Yes	Yes	5.0
P11	pl	Male	31	No	8.0	Yes	Yes	4.0
P12	bc	female	40	Yes	10.0	Yes	Yes	100.0
P13	as	Male	25	No	8.0	Yes	n	0.0
P14	jj	Male	45	No	6.0	Yes	Yes	5.0
P15	al	Male	51	No	10.0	Yes	Yes	5.0
P16	sk	Male	32	No	8.0	Yes	Yes	10.0
P17	jo	Male	31	No	10.0	Yes	Yes	5.0
P18	hk	female	33	No	10.0	Yes	Yes	20.0
P19	mb	Male	37	No	16.0	Yes	Yes	25.0
P20	jk	female	29	No	8.0	Yes	Yes	1.0
P21	ct	Male	33	Yes	10.0	Yes	Yes	8.0
P22	hha	female	36	No	9.0	n	n	0.0
P23	ss	Male	35	Yes	10.0	Yes	Yes	4.0
P24	ma	female	36	Yes	10.0	Yes	Yes	100.0
Responses		23	23	23	24	24	24	24
Tally		15	839	7	224	23	21	357
Result		65.2%	36.5	30.4%	9.3	95.8%	87.5%	14.9
Units		Male	Years	English	Hours per day	Yes	Yes	Messages per day

Topics

- The what, why, and how of empirical research
- Observations and measurements
- Research questions
- Experiment terminology
- Group participation in a real experiment
- Experiment design
- ANOVA statistics and experiment results
- Parts of a research paper



Thank you