MacroEconometric Forecasting

Topic:

Starting analysis

Presented by Munisa Yashnarbekova

Part I – Briefly

What is Empirical Research?

- Empirical research is…
  - observation-based investigation seeking to discover and interpret facts, theories, or laws (relating to humans interacting with computers)

Why do Empirical Research?

- We conduct empirical research to…
  - Answer (and raise!) questions about new or existing user interface designs or interaction techniques
  - Develop or explore models that *describe* or *predict* behaviour (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* †

† *Scientific method* - a body of techniques for investigating phenomena and acquiring new knowledge, as well as for correcting and integrating previous knowledge. It is based on gathering observable, empirical, measurable evidence, subject to the principles of reasoning.(wikipedia)

Part II – The Details (with an HCI context)

Themes

- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

When we do Research, we…

- Observe
- Measure
- Describe
- Compare
- Infer
- Relate
- Predict
- etc.

Empirical

When we do Research, we…

- Observe … human behaviour and response
- Measure … using numbers
- Describe … using numbers
- Compare … using numbers
- Infer … using numbers
- Relate … using numbers
- Predict … using numbers
- etc.

Empirical - capable of being verified or disproved by observation or experiment (Websters dictionary)

So, what is non-empirical research?

Non-Empirical Research

- Non-empirical research (aka qualitative research) is generally concerned with the reasons underlying human behaviour (i.e., the *why* or *how*, as opposed to the *what*, *where*, or *when*)
- Tends to focus on human…
  - thought, feeling, attitude, emotion, reflection,, sentiment, opinion, mood, outlook, manner, approach, strategy, etc.
  - These human qualities are not directly observable or measurable and, therefore, necessitate a different method of inquiry (e.g., case studies, focus groups, cultural probes, personae, etc.)
  - But see… (click here)

Observe

- Observations are gathered…
  - Manually
    - Human observers using log sheets, notebooks, questionnaires, etc.
  - Automatically
    - Sensors, switches, cameras, etc.
    - Computer + software to log events + timestamps

Measure

- A measurement is a recorded observation
- An empirical measurement is a number

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

Kelvin, 1883

Scales of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

crude

sophisticated

Nominal – arbitrary assignment of a code to an attribute, e.g.,
1 = male, 2 = female

Ordinal – rank, e.g.,
1st, 2nd, 3rd, …

Interval – equal distance between units, but no absolute zero point, e.g.,
20° C, 30° C, 40° C, …

Ratio – absolute zero point, therefore ratios are meaningful, e.g.,
20 wpm, 40 wpm, 60 wpm

Use ratio measurements where possible

Ratio Measurements

- Preferred scale of measurement
- With ratio measurements summaries and comparisons are strengthened
- Report “counts” as ratios where possible because they facilitate comparisons
- Example – a 10-word phrase was entered in 30 seconds
  - Bad: *t* = 30 seconds
  - Good: Entry rate = 10 / 0.5 = 20 wpm
- Example – two errors were committed while entering a 10-word (50 character) phrase
  - Bad: *n* = 2 errors
  - Good: Error rate was 2 / 50 = 0.04 = 4%

Observe, Measure… Then What?

- Observations and measurements are gathered in a user study (to get “good” data)
- They, we
  - Describe
  - Compare
  - Infer
  - Relate
  - Predict
  - etc.

These are statistical terms.

Fine, but usually our intent is not statistical.

Our intent is founded on simple well-intentioned “research questions”.

Let’s see…

Themes

- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

Research Questions

- Consider the following questions about a new or existing UI design or interaction technique :
  - Is it viable?
  - Is it as good as or better than current practice?
  - Which of several design alternatives is best?
  - What are its performance limits and capabilities?
  - What are its strengths and weaknesses?
  - Does it work well for novices, for experts?
  - How much practice is required to become proficient?

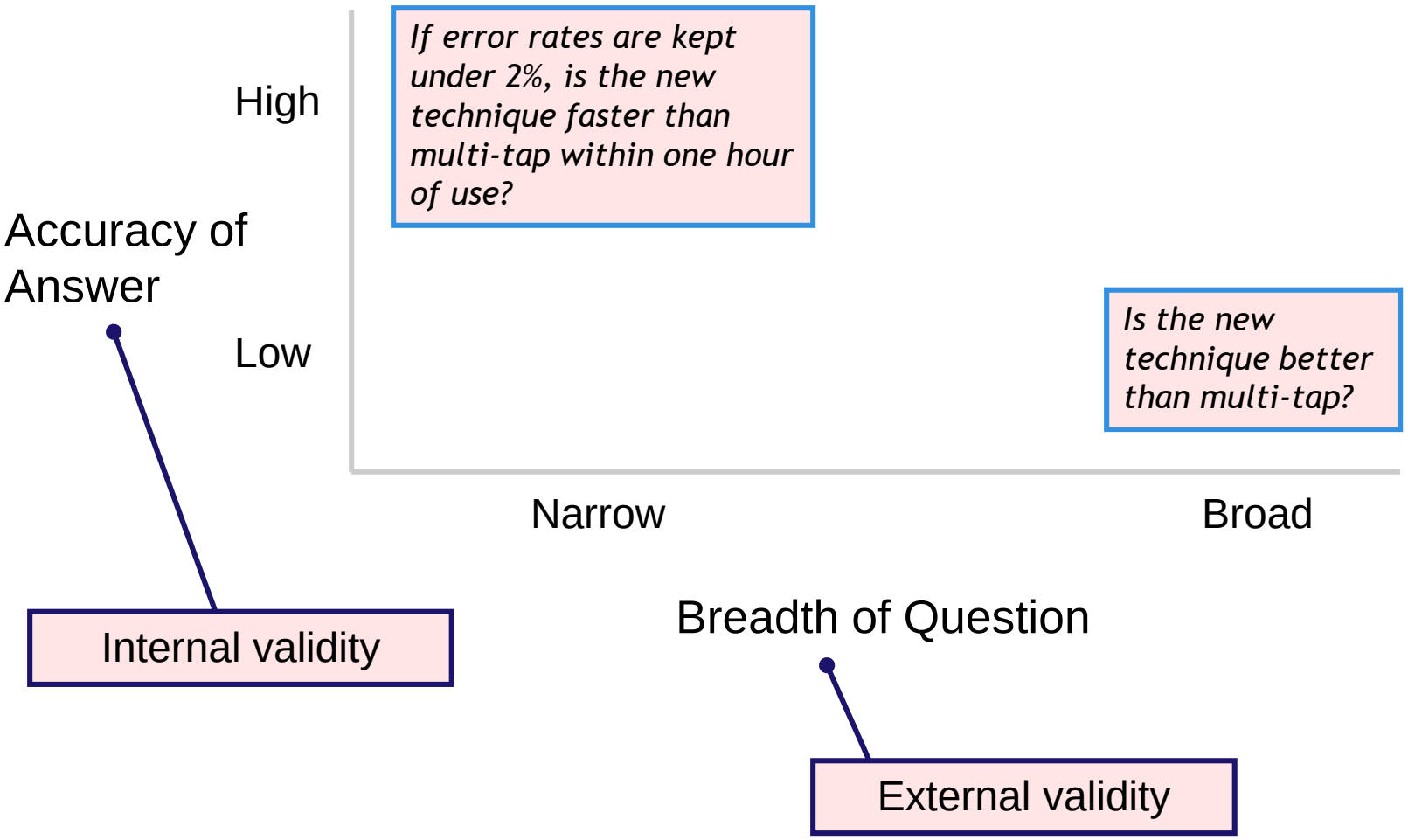
Testable Research Questions

- 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 an idea for a new [technique for entering text on a mobile phone] and you think it's pretty good. In fact, you think it is better than [the commonly used multi-tap technique]. You decide to undertake a program of empirical enquiry to evaluate your idea. What are your research questions?
  - Replace […] as appropriate for other research topics

Research Questions (2)

- Very weak (in an empirical sense)
  - *Is the new technique any good?*
- Weak
  - *Is the new technique better than multi-tap?*
- Better
  - *Is the new technique faster than multi-tap?*
- Better still
  - *Is the new technique faster than multi-tap within one hour of use?*
- Even better
  - *If error rates are kept under 2%, is the new technique faster than multi-tap within one hour of use?*

A Tradeoff


If error rates are kept under 2%, is the new technique faster than multi-tap within one hour of use?
Is the new technique better than multi-tap?
High
Low
Accuracy of Answer
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., multitap vs. new)
- Statistically…
  - 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 and randomly across the test conditions

External Validity

- Definition: The extent to which results are generalizable to other people and other situations
- Statistically…
  - People
    - The participants are representative of the broader intended population of users
  - Situations
    - 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 projection screen (not a CRT)
  - 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 require participants to…
  - Enter representative samples of text (e.g., phrases containing letters, numbers, punctuation, etc.)
  - Edit and correct mistakes as they would normally
- 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

Strive for the Best of Both Worlds

- Internal and external validity are increased by…
  - Posing multiple narrow (***testable***) questions that cover the range of outcomes influencing the broader (***untestable***) questions
    - E.g., a technique that is ***faster***, is ***more accurate***, takes ***fewer steps***, is ***easy to learn***, and is ***easy to remember***, is generally ***better***
- Fortunately…
  - There is usually a positive correlation between the ***testable*** and ***untestable*** questions
    - I.e., participants generally find a UI ***better*** if it is ***faster***, ***more accurate***, ***takes fewer steps***, etc.

One-of vs. Comparative

- Many user studies in HCI are one-of
  - I.e., a new user interface is designed and a user study is conducted to find strengths and weaknesses
- Much better to do a comparative evaluation
  - I.e., A new user interface is designed and it is compared with an alternative design to determine which is better
- The alternative may be
  - A variation in the new design
  - An established design (perhaps a “baseline condition”)
- More than two interfaces may be compared
- Testable research questions are comparative!
- See the paper in *CHI 2006* by Tohidi et al.

Answering Research Questions

- We want to know if the measured performance on a variable (e.g., speed) is different between test conditions, so…
  - We conduct a user study (more on this soon) 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?

Next slide

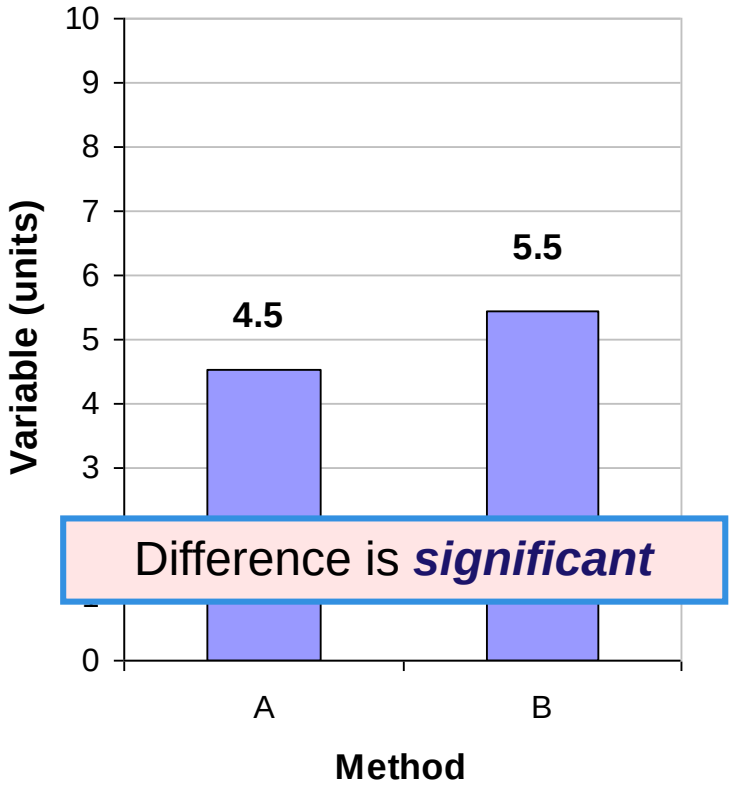
Answering Empirical Questions (2)

- 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?
- Question #1 – obvious (some difference is likely)
- Question #2 – statistics can't help (Is a 5% difference large or small?)
- Question #3 – statistics can help
- Question #4 – statistics can't help (Is a 5% difference useful? People resist change!)
- The basic statistical tool for Question #3 is the analysis of variance (anova)

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?

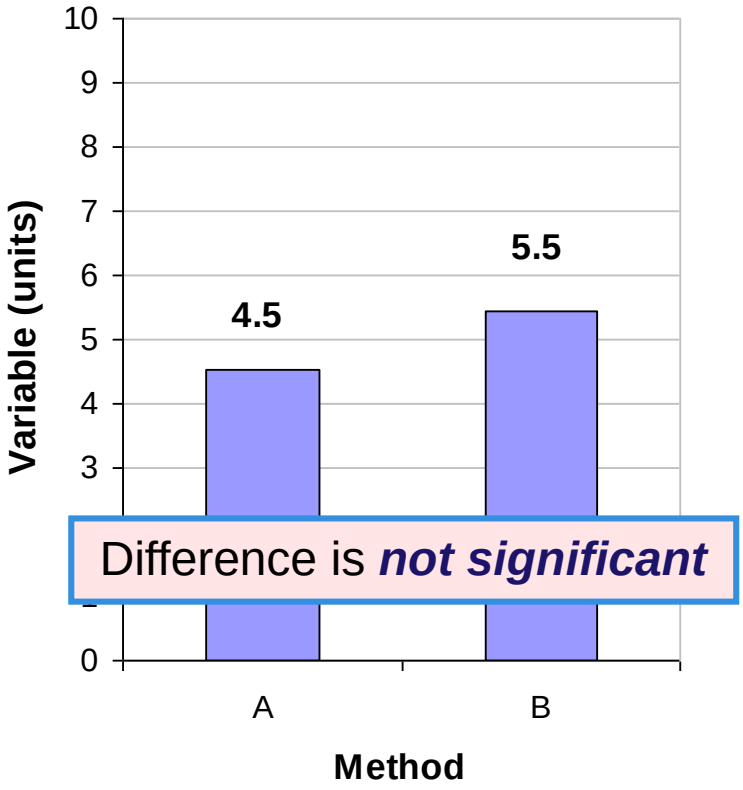
Example #1



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

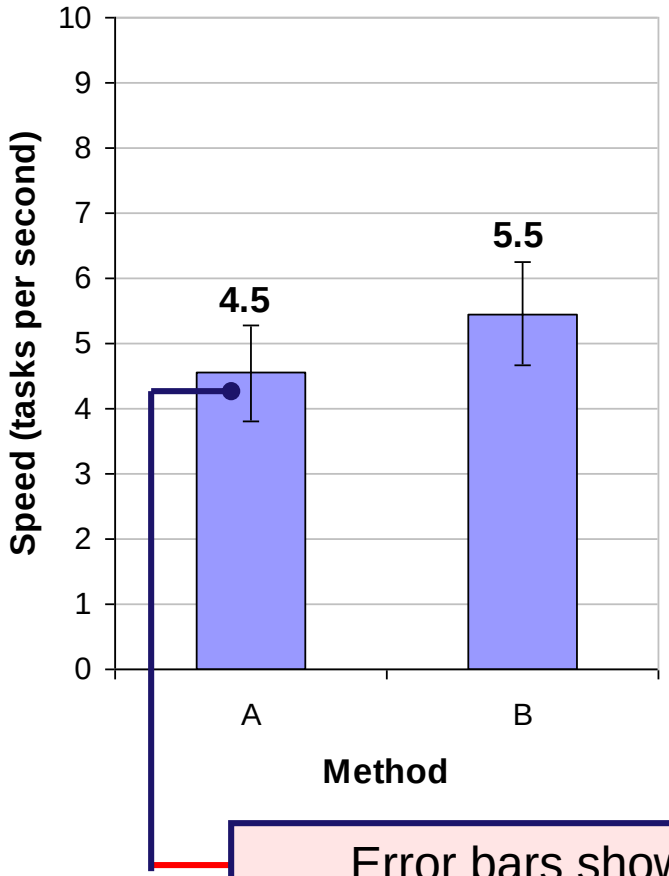


Difference is ***not significant***

"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 that the difference in the means is due to chance

Reported as…

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

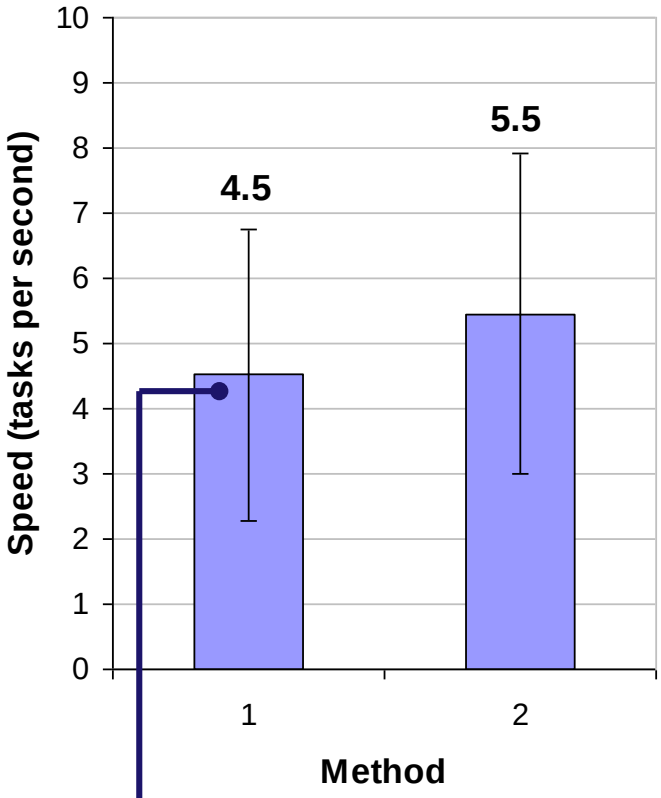
Thresholds for "p"
- .05
- .01
- .005
- .001
- .0005
- .0001

How to Report an *F*-statistic

There was a significant main 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 signn
  - Space after comma
  - Space 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 that the difference in the means is due to chance

Reported as…

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

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

Anova Demo - *StatView*†

Files:

`AnovaExample1.svd`

`AnovaExample2.svd`

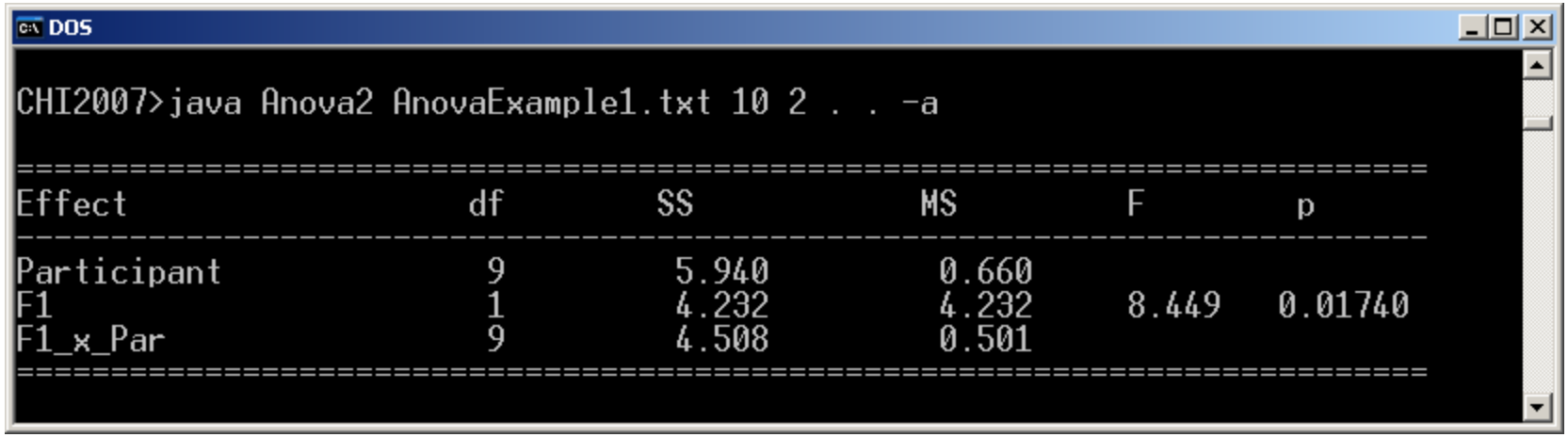
† Now sold as JMP (see http://www.statview.com)

Anova Demo – *Anova2* †

Files:

`AnovaExample1.txt`

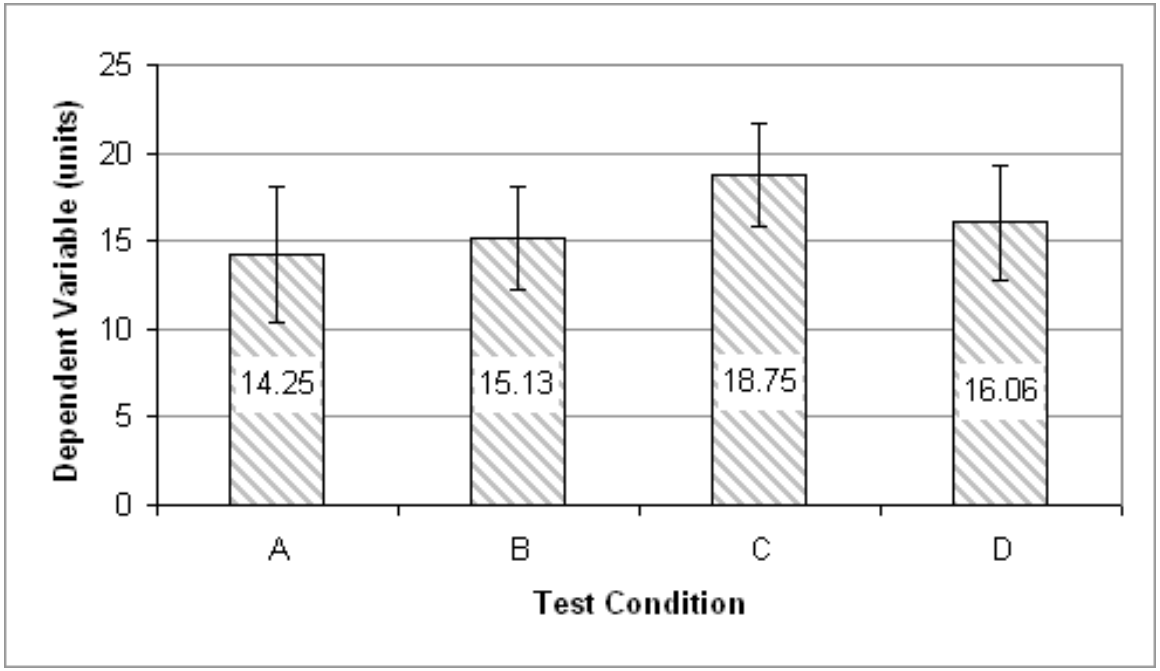
`AnovaExample2.txt`



† This program and its API are available free to attendees of this course. Click here to view API

More Than Two Test Conditions

Participant	Test Condition			
	A	B	C	D
1	11	11	21	16
2	18	11	22	15
3	17	10	18	13
4	19	15	21	20
5	13	17	23	10
6	10	15	15	20
7	14	14	15	13
8	13	14	19	18
9	19	18	16	12
10	10	17	21	18
11	10	19	22	13
12	16	14	18	20
13	10	20	17	19
14	10	13	21	18
15	20	17	14	18
16	18	17	17	14
Mean	14.25	15.13	18.75	16.06
SD	3.84	2.94	2.89	3.23

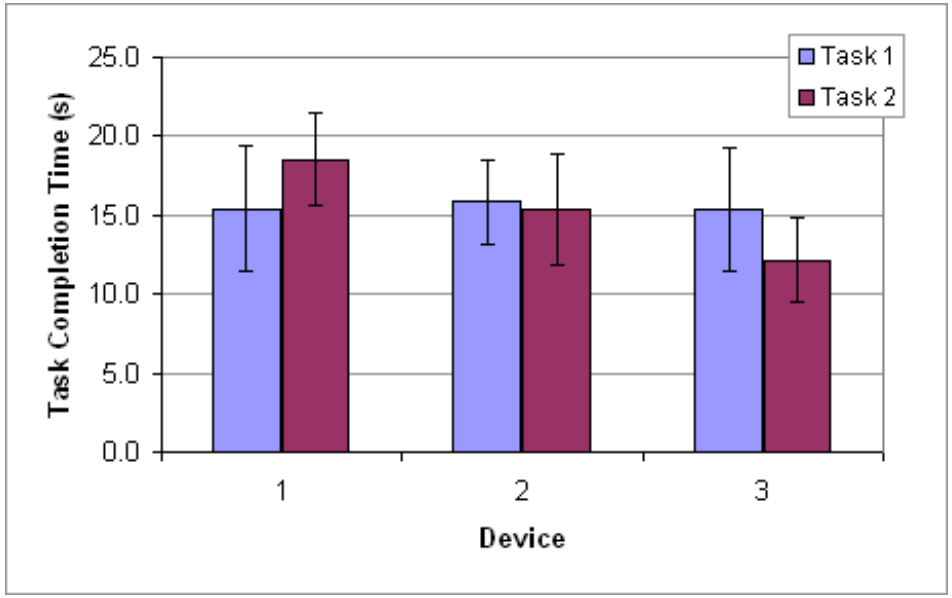


ANOVA Table for Dependent Variable (units)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	15	81.109	5.407				
Test Condition	3	182.172	60.724	4.954	.0047	14.862	.896
Test Condition * Subject	45	551.578	12.257				

Two Factors

Participant	Test Conditions (3 devices x 2 tasks)					
	Device 1		Device 2		Device 3	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
1	11	18	15	13	20	14
2	10	14	17	15	11	13
3	10	23	13	20	20	16
4	18	18	11	12	11	10
5	20	21	19	14	19	8
6	14	21	20	11	17	13
7	14	16	15	20	16	12
8	20	21	18	20	14	12
9	14	15	13	17	16	14
10	20	15	18	10	11	16
11	14	20	15	16	10	9
12	20	20	16	16	20	9
Mean	15.4	18.5	15.8	15.3	15.4	12.2
SD	4.01	2.94	2.69	3.50	3.92	2.69



ANOVA Table for Task Completion Time (s)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	11	134.778	12.253				
Device	2	121.028	60.514	5.865	.0091	11.731	.831
Device * Subject	22	226.972	10.317				
Task	1	.889	.889	.076	.7875	.076	.057
Task * Subject	11	128.111	11.646				
Device * Task	2	121.028	60.514	5.435	.0121	10.869	.798
Device * Task * Subject	22	244.972	11.135				

Themes

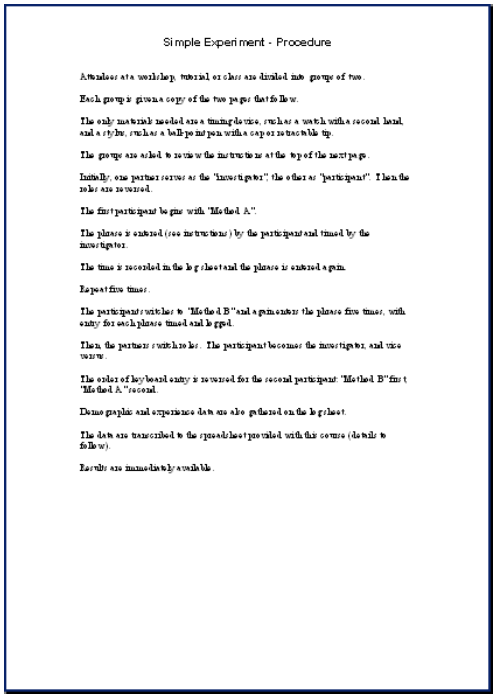
- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

Group Participation†

- 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

†This section may be omitted or shortened depending on the time available

Handout (3 pages)



Simple Experiment - Procedure

Attendees at a workshop, tutorial or class are divided into groups of two.

Each group is given a copy of the two pages that follow.

The only materials needed are a timing device, such as a watch with a second hand, and a stylus, such as a ballpoint pen with a capor retractable tip.

The groups are asked to review the instructions at the top of the next page.

Initially, one partner serves as the "investigator", the other as "participant". Then the roles are reversed.

The first participant begins with "Method A".

The phrase is entered (see instructions) by the participant and timed by the investigator.

The time is recorded in the log sheet and the phrase is entered again.

Repeat five times.

The participants switches to "Method B" and again enters the phrase five times, with entry for each phrase timed and logged.

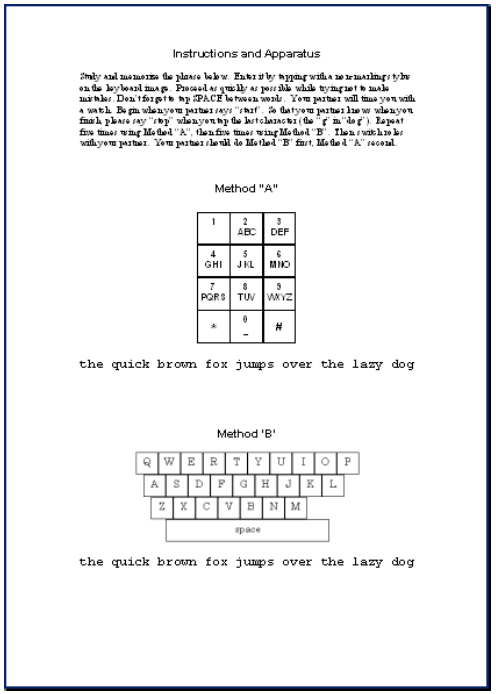
Then, the partners switch roles. The participant becomes the investigator, and vice versa.

The order of keyboard entry is reversed for the second participant: "Method B" first, "Method A." second.

Demographic and experience data are also gathered on the log sheet.

The data are transcribed to the spreadsheet provided with this course (details to follow).

Results are immediately available.



Instructions and Apparatus

Study and memorize the phrase below. Enter it by tapping with a no.mal finger/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 shall do Method "B" first, Method "A" second.

Method "A"

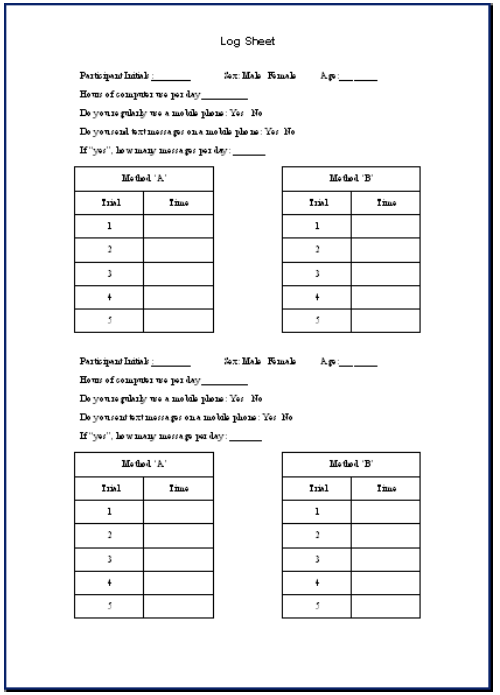
1	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PQRS	8 TUV	9 WXYZ
*	0 _	#

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: ___ ____

Hours of computer use per day: _____

Do you use gularly 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'	
Trial	Time
1	
2	
3	
4	
5	

Method 'B'	
Trial	Time
1	
2	
3	
4	
5	

Participant Initials: _____ Sex: Male Female Age: ___ ____

Hours of computer use per day: _____

Do you use gularly 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'	
Trial	Time
1	
2	
3	
4	
5	

Method 'B'	
Trial	Time
1	
2	
3	
4	
5	

Full-size copies of the handout pages will be distributed during the course. Click here to view.

Do the Experiment

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

Results

Review

When we do Empirical Research, we…

- Observe
- Measure
- Describe
- Compare
- Infer
- Relate
- Predict
- etc.

Entry Time (seconds)												
Participant	Initials	Phone (Method A)					Qwerty (Method B)					Group
		1	2	3	4	5	1	2	3	4	5	
P1	SH	21.0	21.0	24.0	17.0	17.0	21.0	20.0	20.0	17.0	16.0	1
P2	PV	28.0	24.0	25.0	24.0	22.0	18.0	15.0	15.0	16.0	15.0	1
P3	EV	27.0	25.0	22.0	23.0	23.0	20.0	14.0	14.0	15.0	13.0	1
P4	MM	25.0	24.0	19.0	25.0	20.0	21.0	27.0	18.0	20.0	18.0	1
P5	HV	30.0	28.0	24.0				22.0	20.0	23.0	22.0	1
P6	OM	45.0	37.0	43.0				20.0	20.0	18.0	17.0	1
P7	TV	28.0	25.0	23.0	21.0	21.0	17.0	13.0	14.0	12.0	12.0	1
P8	LM	28.0	20.0	19.0	24.0	19.0	23.0	20.0	19.0	18.0	18.0	1
P9	MM	25.0	25.0	21.0				17.0	14.0	15.0	14.0	1
P10	SU	20.0	18.0	17.0				18.0	15.0	16.0	13.0	1
P11	JR	28.0	23.0	21.0	19.0	18.0	20.0	18.0	19.0	16.0	16.0	1
P42	AH	36.0	37.0	28.0	27.0	26.0	33.0	29.0	25.0	23.0	24.0	2
P43	PP	35.0	29.0	28.0	27.0	24.0	22.0	20.0	18.0	18.0	16.0	2
P44	KT	53.0	46.0	43.0	36.0	35.0	29.0	23.0	22.0	18.0	17.0	2
P45	DH	30.0	28.0	27.0	25.0	26.0	22.0	20.0	18.0	16.0	16.0	2
P46	RV	24.0	20.0	20.0	15.0	13.0	23.0	19.0	19.0	17.0	18.0	2
P47	SK	30.0	30.0	24.0	16.0	26.0	22.0	17.0	20.0	16.0	15.0	2
P48	RH	38.0	34.0	32.0	32.0	27.0	27.0	19.0	17.0	16.0	13.0	2

Observe

Measure

. . .

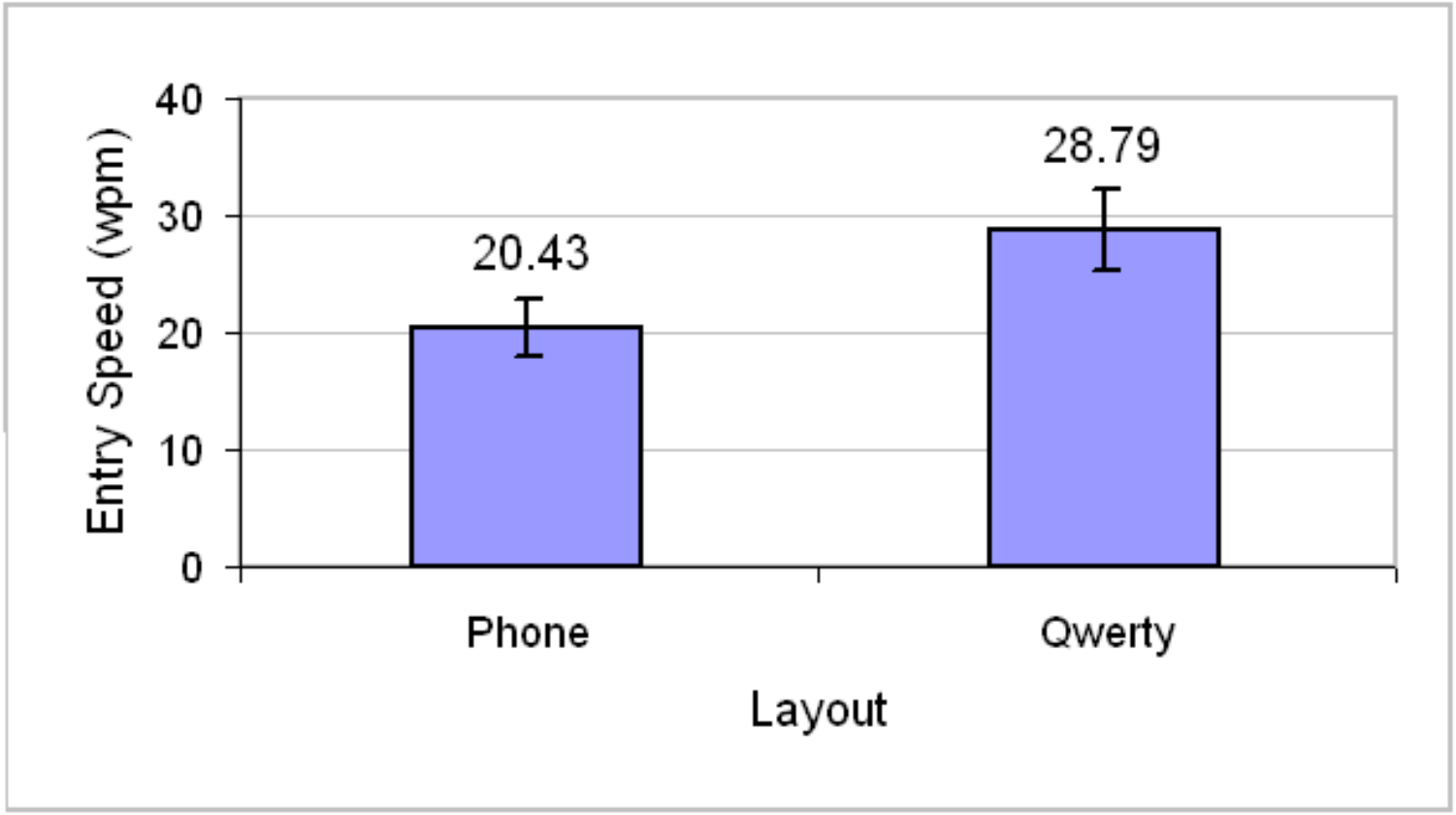
SimpleExperiment-results-TAUCHI-02-2007.xls

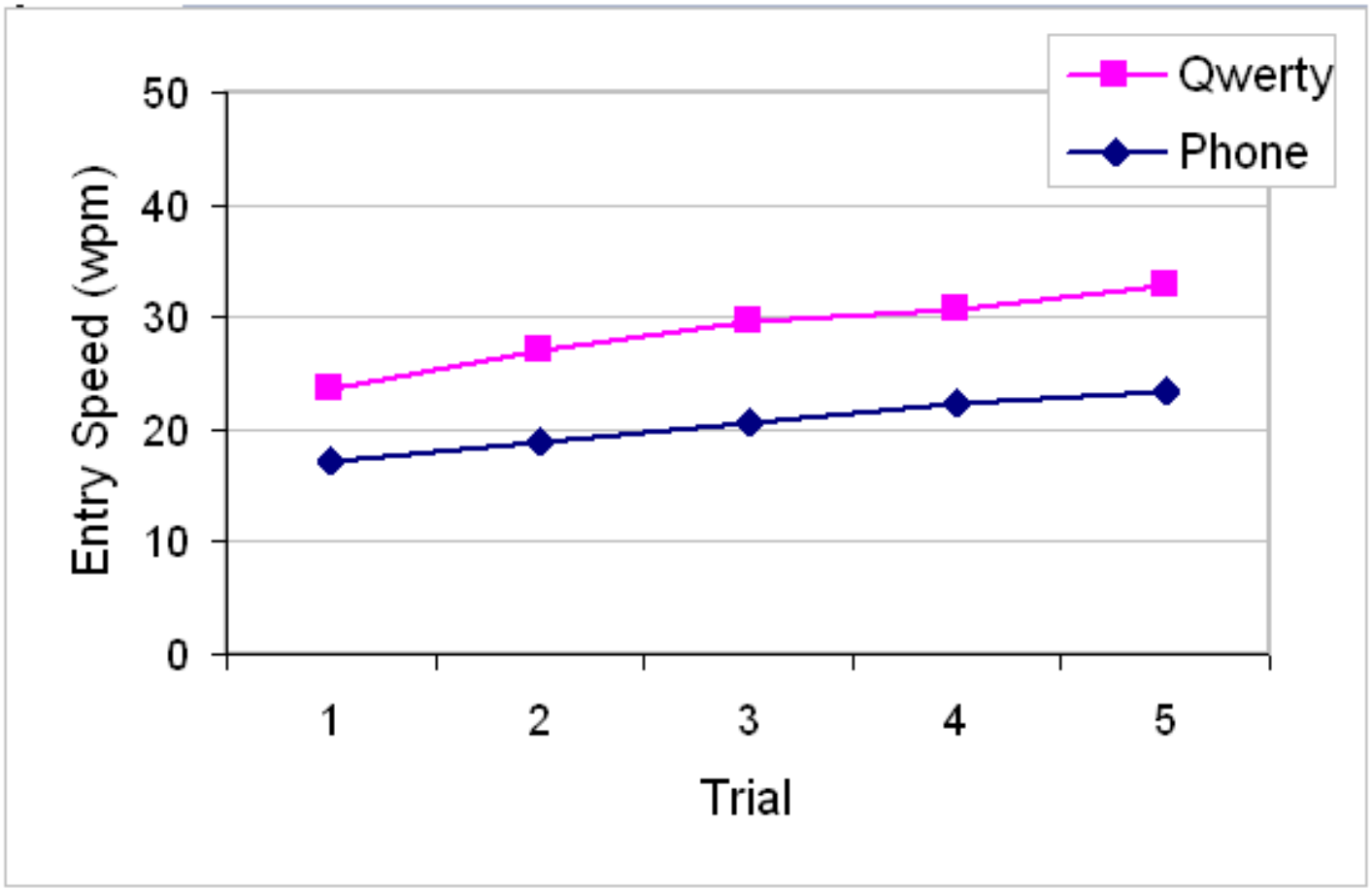
Entry Speed (wpm)												
Participant	Initials	Phone (Method A)					Qwerty (Method B)					Group
		1	2	3	4	5	1	2	3	4	5	
P2	SH	24.57	24.57	21.50	30.35	30.35	24.57	25.80	25.80	30.35	32.25	1
P3	PV	18.43	21.50	20.64	21.50	23.45	28.67	34.40	34.40	32.25	34.40	1
P4	EV	19.11	20.64	23.45	22.43	22.43	25.80	36.86	36.86	34.40	39.69	1
P5	MM	20.64	21.50	27.16	20.64	25.80	24.57	19.11	28.67	25.80	28.67	1
P6	HV	17.20	18.43	21.50	19.85	19.85	19.85	23.45	25.80	22.43	23.45	1
P37	HN	27.16	25.80	24.57	27.16	27.16	25.80	23.45	30.35	32.25	30.35	2
P38	AA	18.43	19.85	25.80	27.16	25.80	27.16	34.40	34.40	34.40	32.25	2
P39	JL	15.18	12.59	17.20	20.64	22.43	19.85	30.35	36.86	36.86	51.60	2
P40	NH	17.20	19.11	22.43	24.57	25.80	20.64	24.57	25.80	25.80	24.57	2
P41	SK	12.90	14.33	13.58	16.65	17.79	21.50	20.64	23.45	24.57	24.57	2
P42	AI	6.37	7.27	7.37	7.37	7.82	17.79	21.50	22.43	25.80	28.67	2
P43	AH				19.11	19.85				2.43	21.50	2
P44	PP				19.11	21.50				3.67	32.25	2
P45	KT	9.74	11.22	12.00	14.33	14.74	17.79	22.43	23.45	28.67	30.35	2
P46	DH	17.20	18.43	19.11	20.64	19.85	23.45	25.80	28.67	32.25	32.25	2
P47	RV	21.50	25.80	25.80	34.40	39.69	22.43	27.16	27.16	30.35	28.67	2
P48	SK	17.20	17.20	21.50	32.25	19.85	23.45	30.35	25.80	32.25	34.40	2
P49	RH	13.58	15.18	16.13	16.13	19.11	19.11	27.16	30.35	32.25	39.69	2
	Mean	17.08	18.91	20.53	22.23	23.37	23.65	27.04	29.68	30.79	32.80	*Mean*
	SD	4.59	4.93	5.28	6.46	6.24	4.75	5.28	6.07	5.41	6.48	*SD*
					Min	6.37				*Min*	12.59	*Min*
					Max	39.69				*Max*	51.60	*Max*

Describe

Compare

Method	Trial 1	2	3	4	5	Mean	SD	
Phone	17.08	18.91	20.53	22.23	23.37	20.43	2.52	
Qwerty	23.65	27.04	29.68	30.79	32.80	28.79	3.55	41.0% faster
Improvement		10.7%	20.2%	30.2%	36.9%			
Improvement		14.3%	25.5%	30.2%	38.7%			





Note: Use bar chart for nominal data (previous slide), line chart for continuous data (above)

Infer

ANOVA Table for Entry Speed (wpm)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	47	9386.065	199.704				
Keyboard	1	8396.826	8396.826	131.150	<.0001	131.150	1.000
Keyboard * Subject	47	3009.156	64.025				
Trial	4	3518.713	879.678	134.193	<.0001	536.773	1.000
Trial * Subject	188	1232.397	6.555				
Keyboard * Trial	4	121.482	30.371	5.389	.0004	21.555	.979
Keyboard * Trial * Subject	188	1059.535	5.636				

There was a significant effect of keyboard layout on entry speed ($F_{1,47} = 131.2$, $p < .0001$).

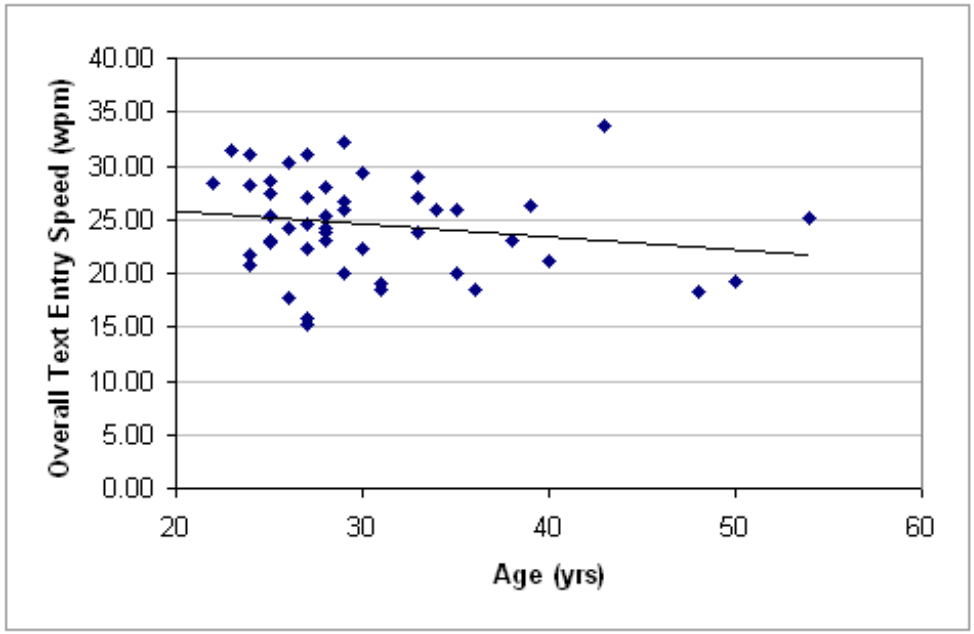
Questionnaire							
Participant	Initials	Sex	Age	Hours of computer use per day?	Do you regularly use a mobile phone?	Do you send text messages on a mobile phone?	If yes, how many messages per day?
P1	SH	f	27	9	yes	yes	2.0
P2	PV	m	33	8	yes	yes	1.0
P3	EV	f	24	6	yes	yes	0.5
P4	MM	f	26	9	yes	yes	2.0
P5	HV	f	40	8	yes	yes	10.0
P6	OM	f	50	5	yes	yes	5.0
P7	TV	m	26	10	yes	yes	2.0
P8	LM	f	28	8	yes	yes	5.0
P9	MM	f	25	8	yes	yes	0.5
P10	SU	f	23	8	yes	yes	15.0
P43	MK	f	39	7	yes	yes	3.0
P41	AI	m	27	9	yes	yes	1.0
P42	AH	f	48	5	yes	yes	3.5
P43	PP	m	38	7	yes	yes	10.0
P44	KT	m	36	10	yes	yes	1.0
P45	DH	m	28	13	yes	yes	1.0
P46	RV	f	22	8	yes	yes	10.0
P47	SK	f	25	6	yes	yes	2.0
P48	RH	m	25	4	yes	yes	1.0
Responses		48	47	48	48	48	48
Tally		19	1435	370	48	48	190.0
Result		**39.6%**	**30.5**	**7.7**	**100.0%**	**100.0%**	**4.0**
(units)		**male**	**years**	**hours**	**yes**	**yes**	**messages/day**

Relate

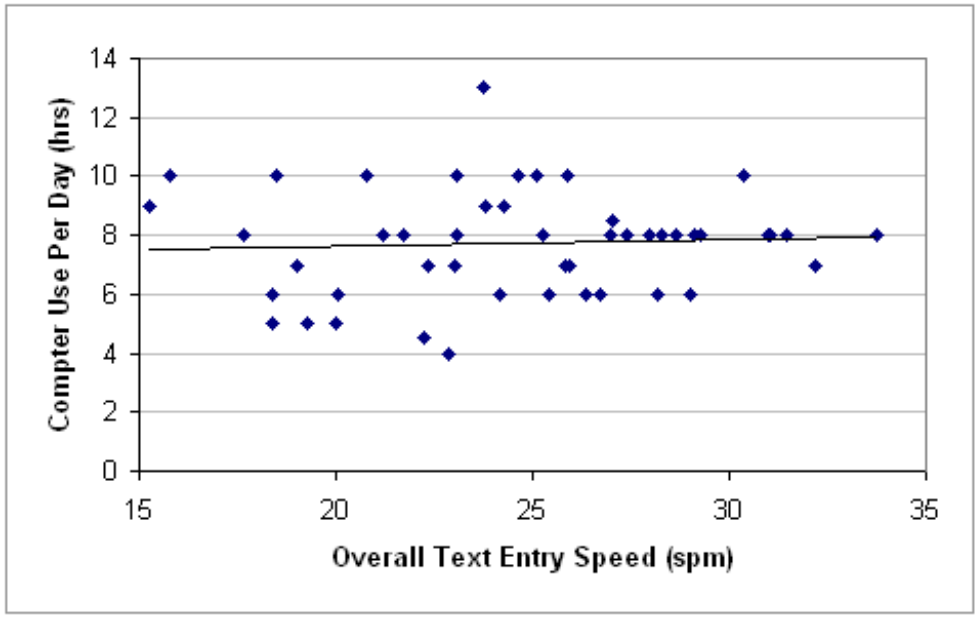
...

?

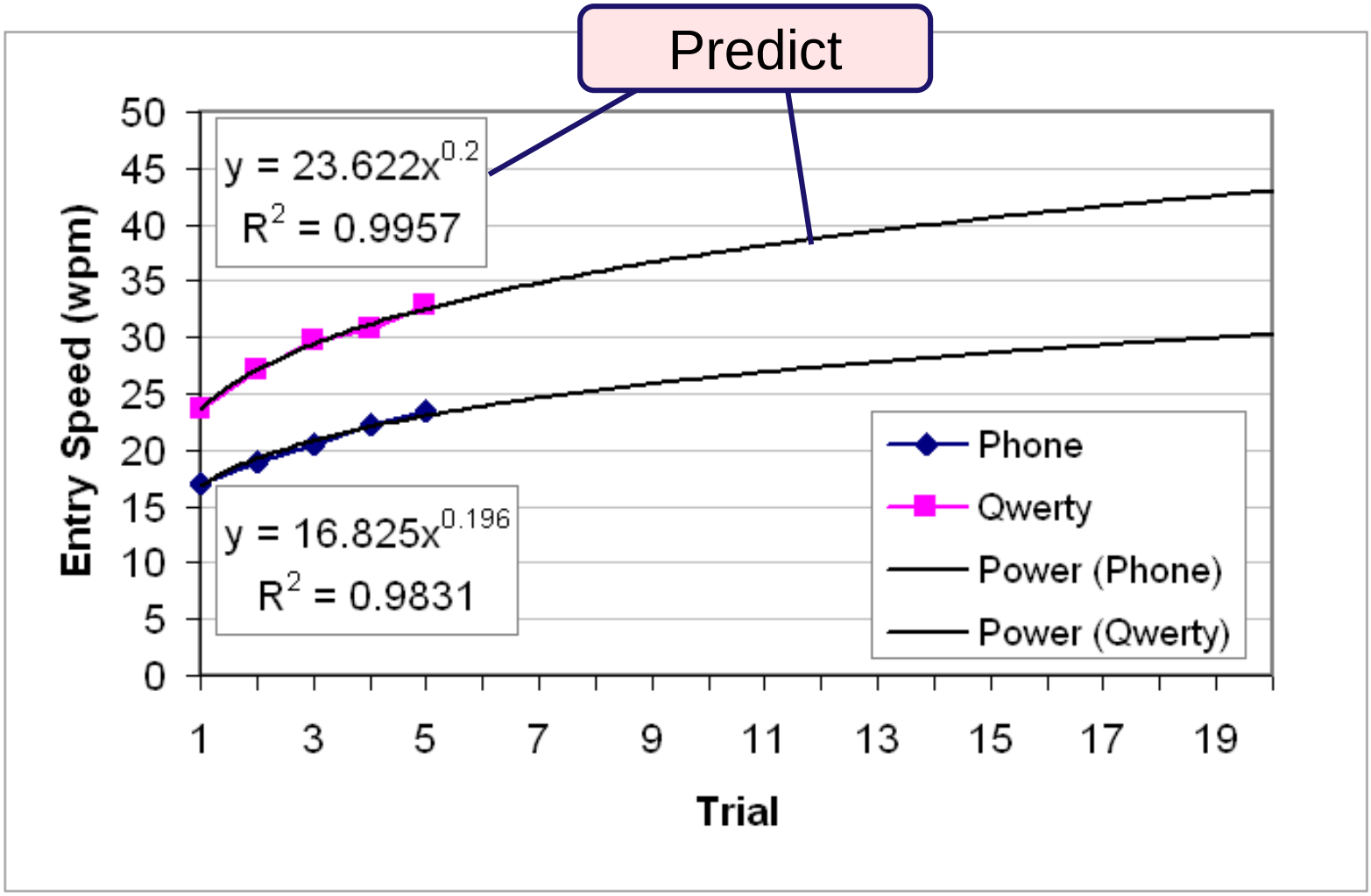
Next slide



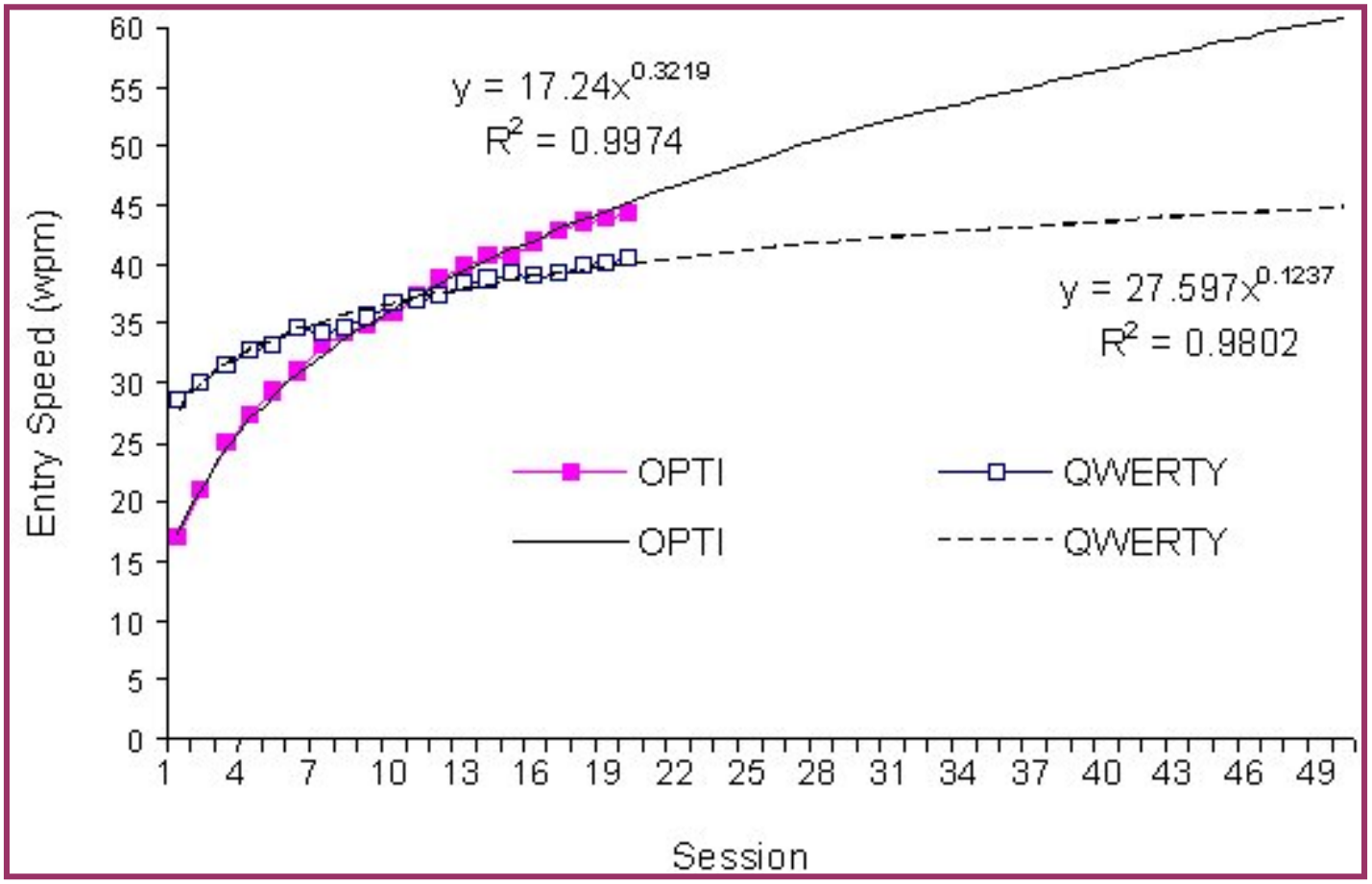
Correlation
r =
-0.1325204



Correlation
r =
0.0563545


Predict
y = 23.622x^0.2
R^2 = 0.9957
y = 16.825x^0.196
R^2 = 0.9831
Phone
Qwerty
Power (Phone)
Power (Qwerty)
Entry Speed (wpm)
Trial
0
5
10
15
20
25
30
35
40
45
50
1
3
5
7
9
11
13
15
17
19

Prediction Equation from a Longitudinal Study



Click here to view paper.

Themes

- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

Experiment Design

- Experiment design is the process of deciding which variables to use, what tasks and procedure to use, how many participants to use and how to solicit them, and so on
- Let’s work on the terminology…

Experiment Design - Terminology

- Terms to know
  - Participant
  - Independent variable (test conditions)
  - Dependent variable
  - 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
- Previously the term subjects was used, but it is no longer in vogue
- When referring specifically to the experiment, use the term participants (e.g., “all participants exhibited a high error rate…”)
- General comments on the problem or conclusions drawn 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

How Many Participants?

- The Answer: It depends!
- Too many:
  - Results are statistically significant, even where the differences are miniscule and of no practical relevance
- Too few:
  - Results are not statistically significant (because of the small sample size), even though there maybe a significant difference in the test conditions (“significant” in the practical sense)
- Guideline:
  - Use approximately the same number of participants as in other similar research †

† Martin, D. W. (2004). *Doing psychology experiments* (6th ed.). Belmont, CA: Wadsworth.

Independent Variable

- An independent variable is a variable that is manipulated through the design of the experiment
- It is “independent” because it is independent of participant behaviour (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 a variable representing the measurements or observations on a independent variable
- Examples include task completion time, speed, accuracy, error rate, throughput, target re-entries, retries, key actions, 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”)

Three “Other” Variables

- Important but usually given less attention are
  - Control variables
  - Random variables
  - Confounding variables

Control Variable

- Circumstances or factors that (a) might influence a dependent variable, but (b) are not under investigation need to be accommodated in some manner
- One way is to control them – to treat them as control variables
- E.g., room lighting, background noise, temperature
- The disadvantage to having too many control variables is that the experiment becomes less generalizable (i.e., less applicable to other situations)

Random Variable

- Instead of controlling all circumstances or factors, some might be allowed to vary randomly
- Such circumstances are random variables
- More variability is introduced in the measures (that's bad!), but the results are more generalizable (that's good!)

Confounding Variable

- Any variable that varies systematically with an independent variable is a confounding variable
- Example 1 – three techniques are compared (A, B, C)
  - All participants are tested on A, followed by B, followed by C
  - Performance might improve due to practice
  - “Practice” is a confounding variable (because it varies systematically with “technique”)
- Example 2 – two search engine interfaces are compared (Google vs. new)
  - All participants have prior experience with Google, but no experience with the new interface
  - “Prior experience” is a confounding variable

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 condition.
- The terms repeated measures and within subjects are synonymous.

Within vs. Between Subjects

- Question: Is it best to assign a factor 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. In this case, there is a tradeoff
- 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. This is bad news.
- 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

Random Order

- Counterbalancing using a Latin Square requires *m* participants, where *m* % *n* = 0 (*n* is the number of test conditions)
- Sometimes this is not practical or possible (e.g., number of participants is unknown)
- Alternatively, learning effects may be minimized by
  - Randomizing the order of presentation
  - Using “all possible orders”; e.g., *n* = 3 →

ABC

ACB

BAC

BCA

CAB

CBA

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 test conditions
- Note: A mixed design is also possible
  - In this case, 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).
  - Click here for an example of a mixed design

Themes

- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

Steps in Empirical Research (1)

Phase I – The Prototype

Think, Analyse, Model, Create, Choose, etc.

Build Prototype

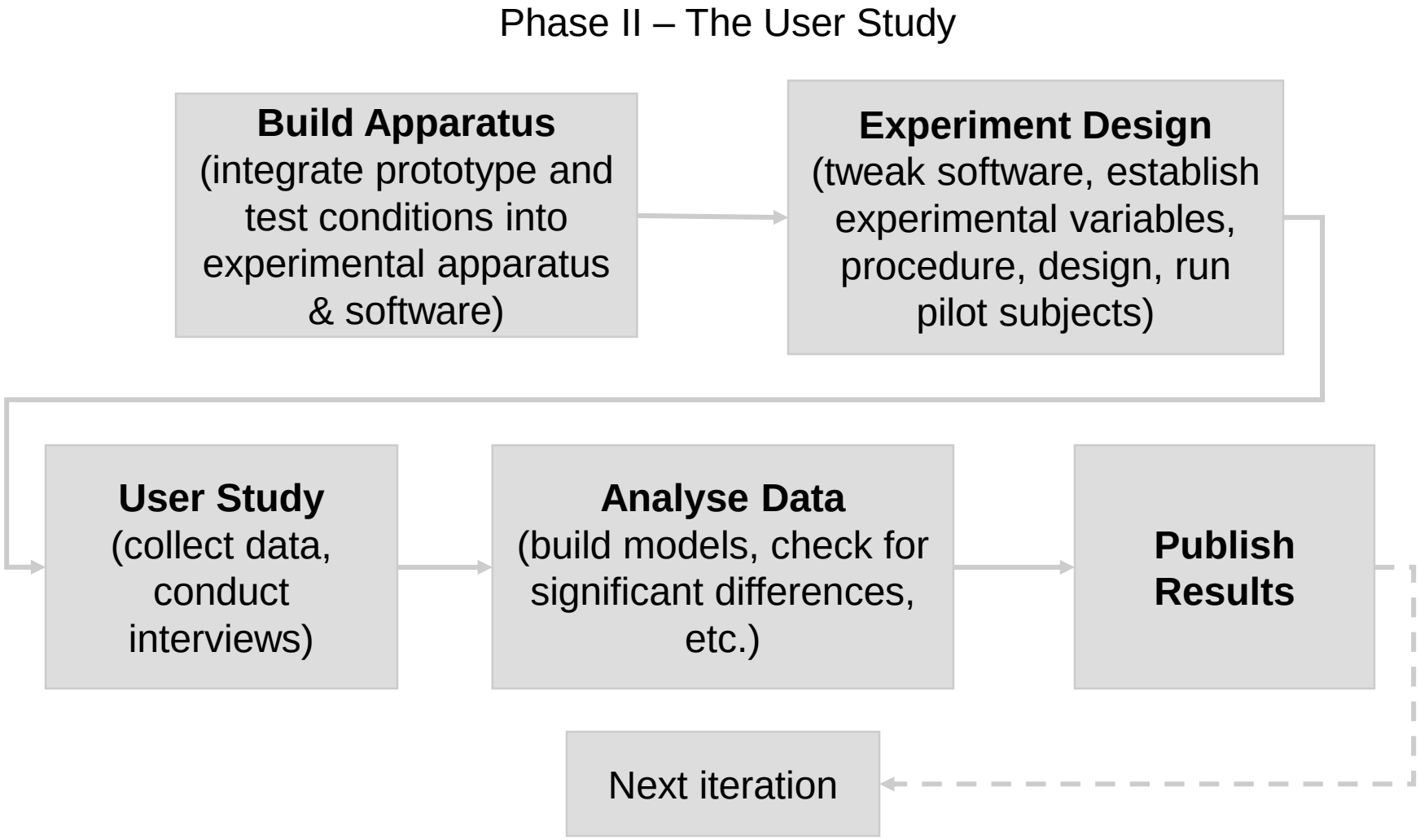
Test, Measure, Compare

Short paper, Poster, Abstract

Iterations are frequent, unstructured, intuitive, informed, …

Research questions “take shape” (I.e., certain measurable aspects of the interaction suggest “test conditions”, and “tasks” for empirical inquiry.

Steps in Empirical Research (2)



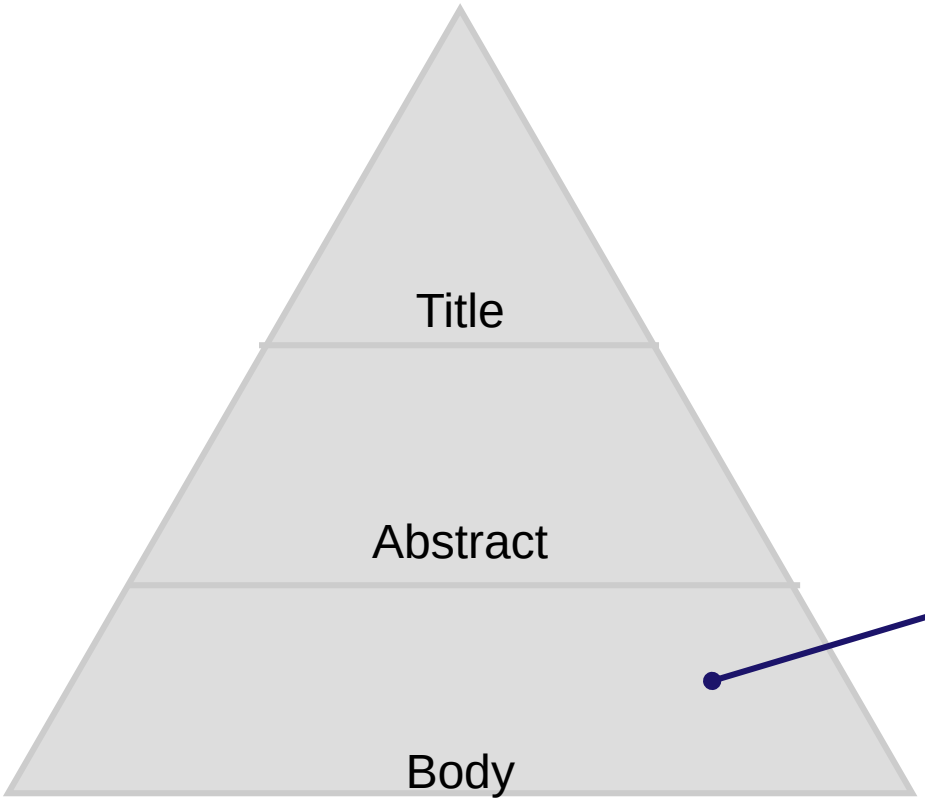
Themes

- Observe and measure
- Research questions
- User studies – group participation
- User studies – terminology
- User studies – step by step summary
- Parts of a research paper

Research Paper

- The final step
- Research is not finished until the results are published!

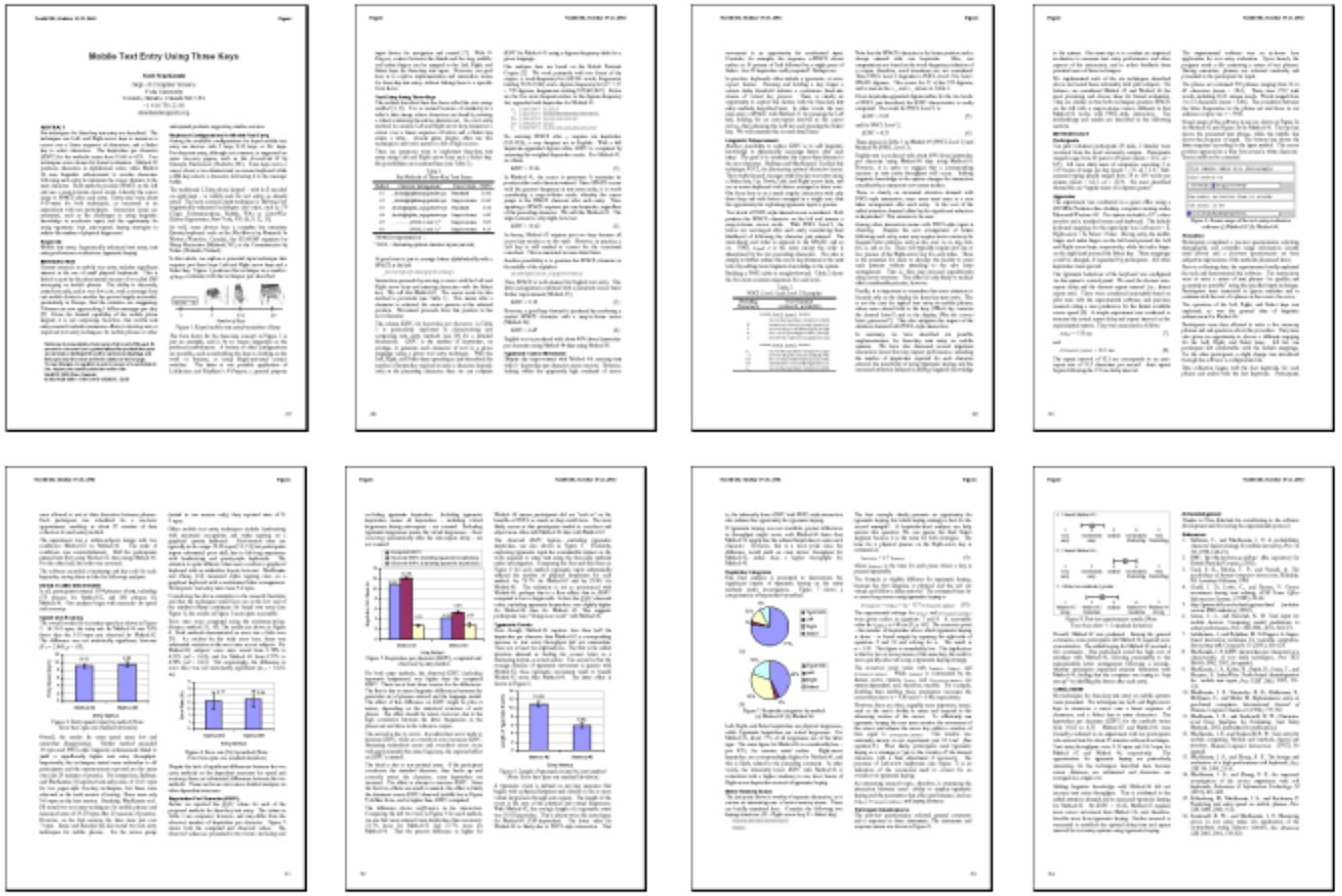
Organization of a Research Paper



Main sections…
- Introduction
- Method
  - Participants
  - Apparatus
  - Procedure
  - Design
- Results and Discussion
- Conclusions

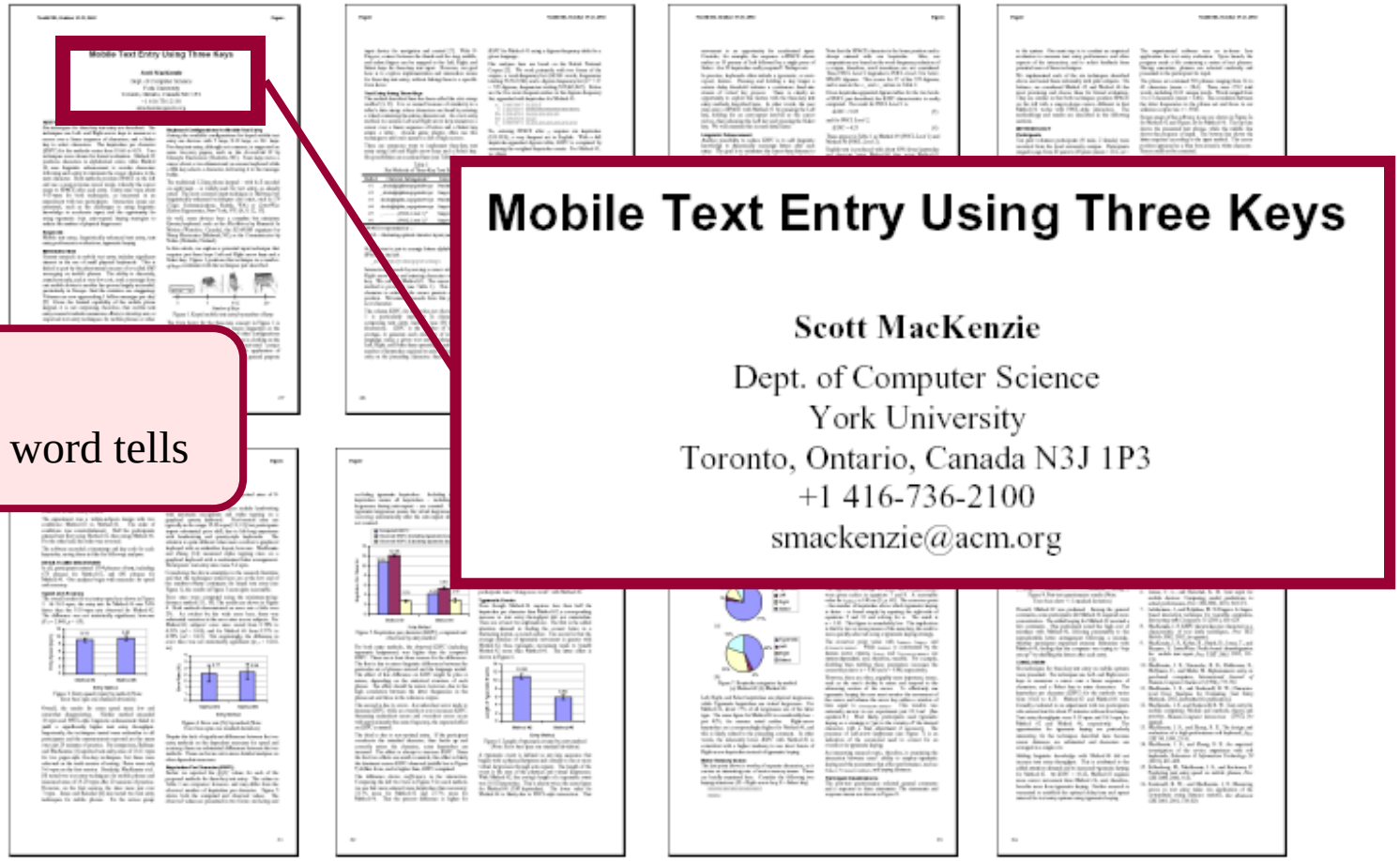
Formatted according to submission requirements of conference or journal (e.g., click here view template for CHI submissions).

Example Publication†



† MacKenzie, I. S. (2002). Mobile text entry using three keys. *Proceedings of NordiCHI 2002*, 27-34. New York: ACM.

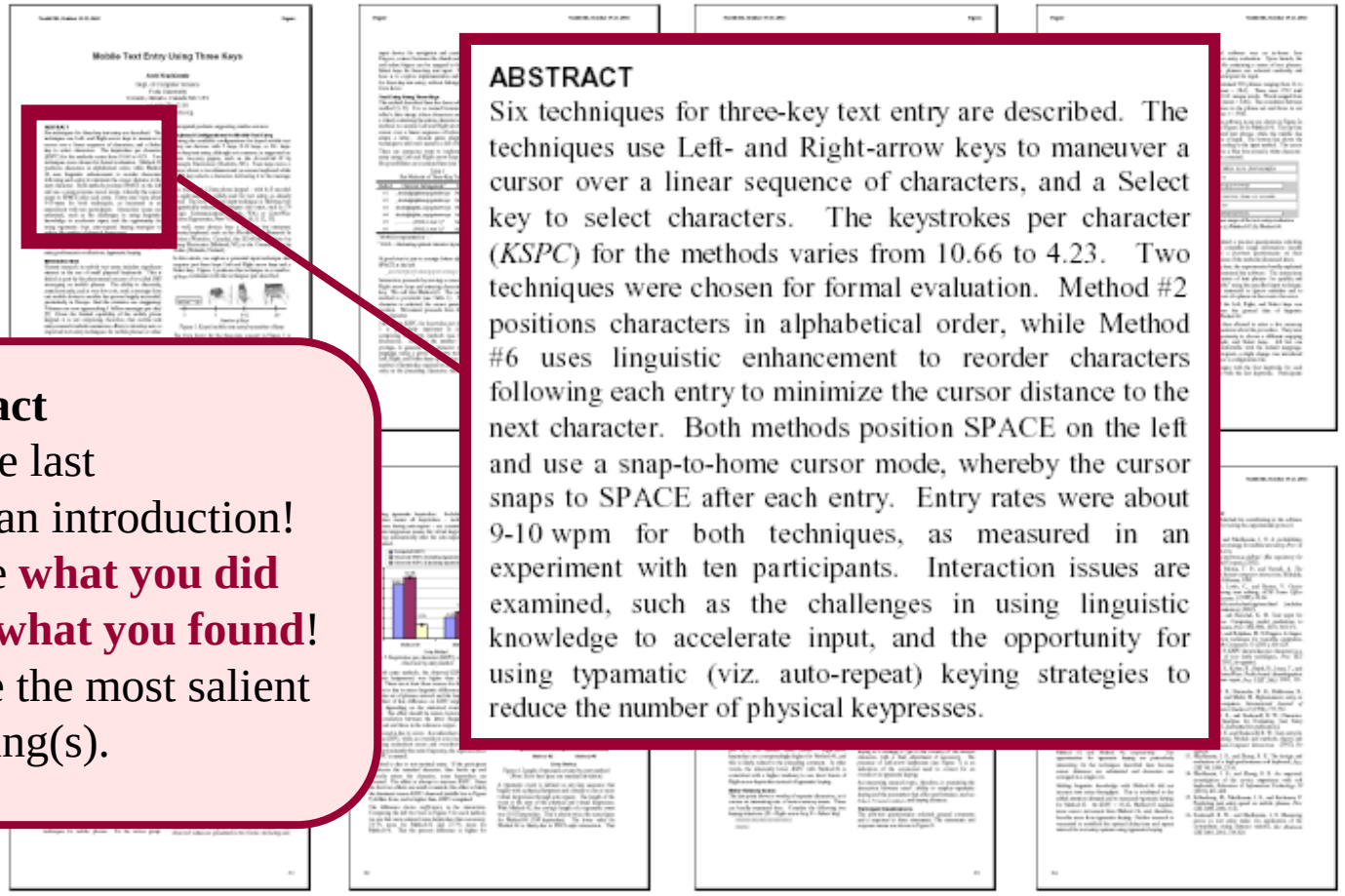
Title, Author(s), Affiliation(s)



Title

- Every word tells

Abstract



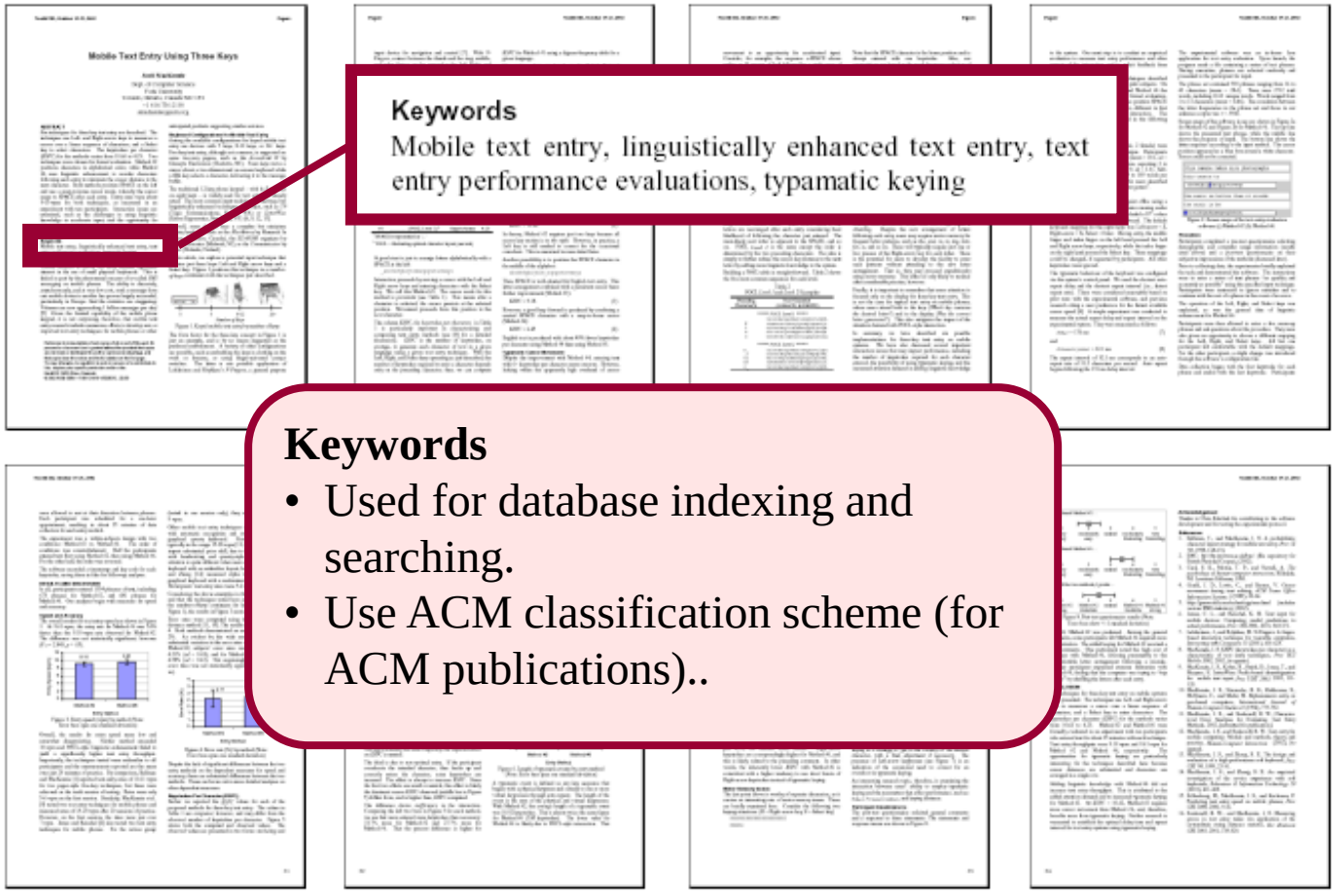
ABSTRACT

Six techniques for three-key text entry are described. The techniques use Left- and Right-arrow keys to maneuver a cursor over a linear sequence of characters, and a Select key to select characters. The keystrokes per character (*KSPC*) for the methods varies from 10.66 to 4.23. Two techniques were chosen for formal evaluation. Method #2 positions characters in alphabetical order, while Method #6 uses linguistic enhancement to reorder characters following each entry to minimize the cursor distance to the next character. Both methods position SPACE on the left and use a snap-to-home cursor mode, whereby the cursor snaps to SPACE after each entry. Entry rates were about 9-10 wpm for both techniques, as measured in an experiment with ten participants. Interaction issues are examined, such as the challenges in using linguistic knowledge to accelerate input, and the opportunity for using typamatic (viz. auto-repeat) keying strategies to reduce the number of physical keypresses.

Abstract

- Write last
- Not an introduction!
- State **what you did** and **what you found**!
- Give the most salient finding(s).

Keywords

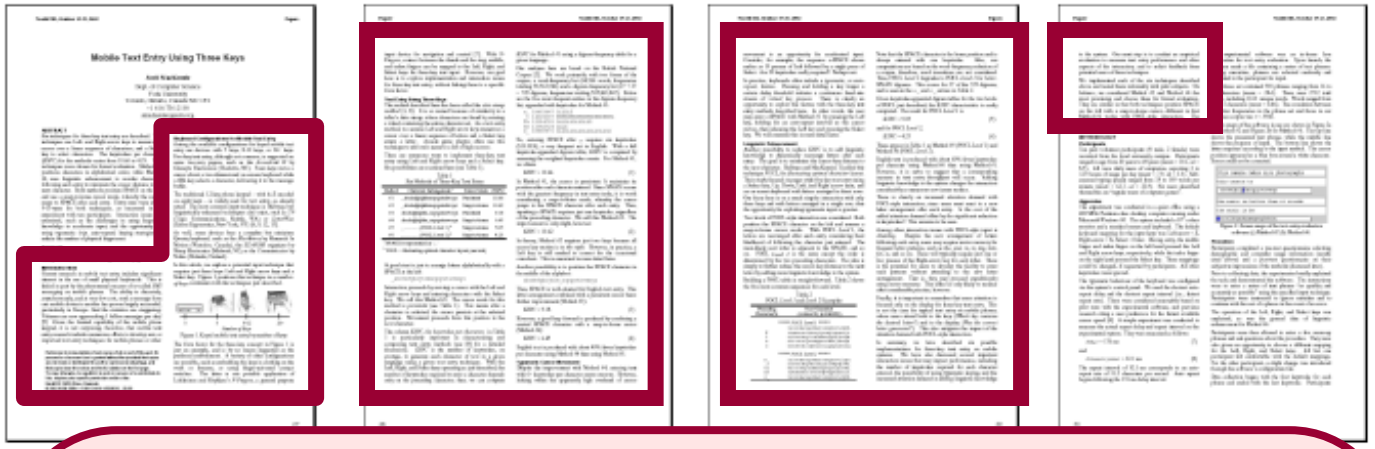


Keywords
Mobile text entry, linguistically enhanced text entry, text entry performance evaluations, typamatic keying

Keywords

- Used for database indexing and searching.
- Use ACM classification scheme (for ACM publications)..

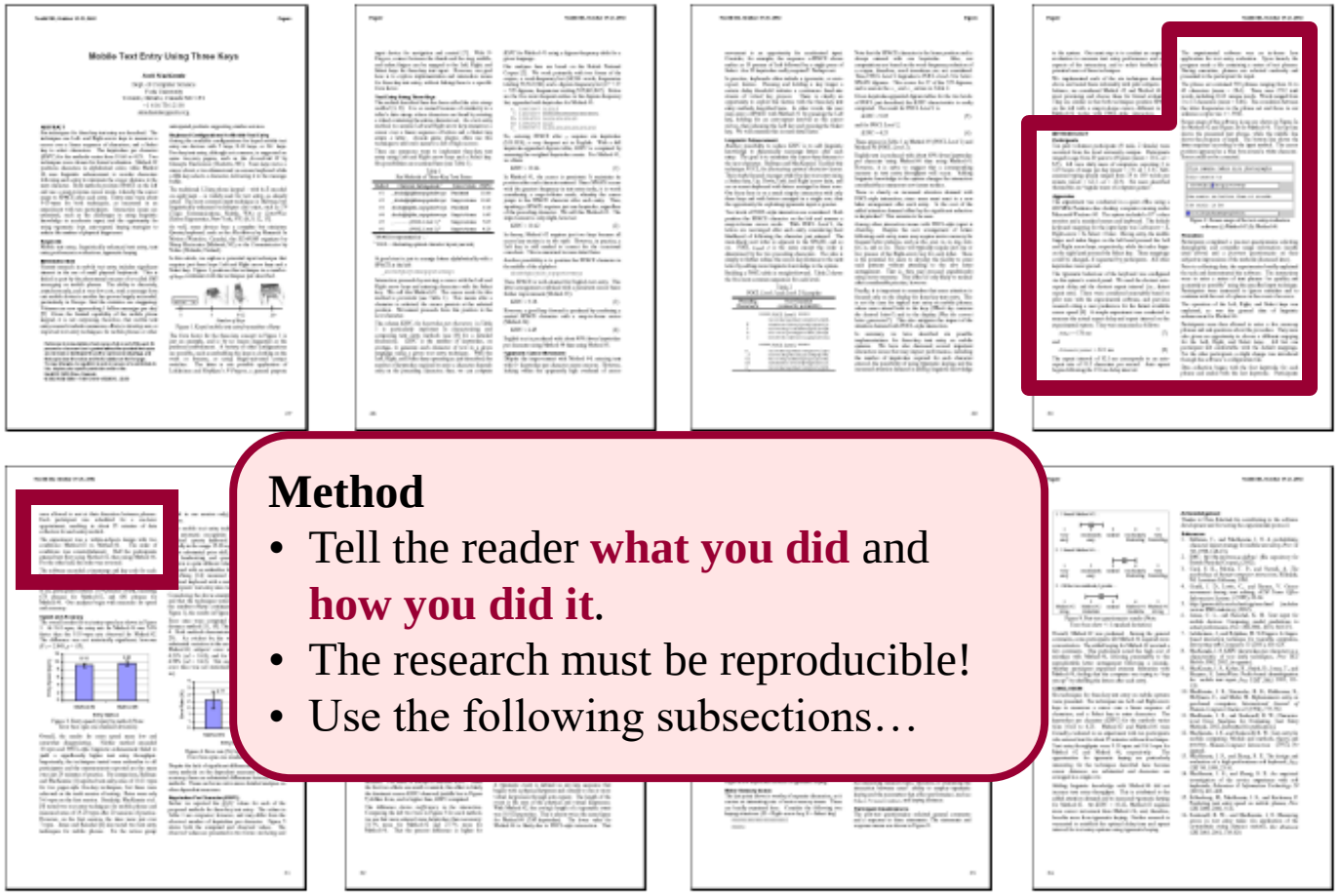
Introduction



Introduction

- Give the context for the research, stating why it is interesting and relevant.
- Identify a UI problem or challenge as it currently exists.
- Give an overview of the contents of the entire paper.
- Identify, describe, cite related work.
- Describe and justify your approach to the problem.
- Follow the formatting requirements of conference or journal.
- **It's your story to tell!**

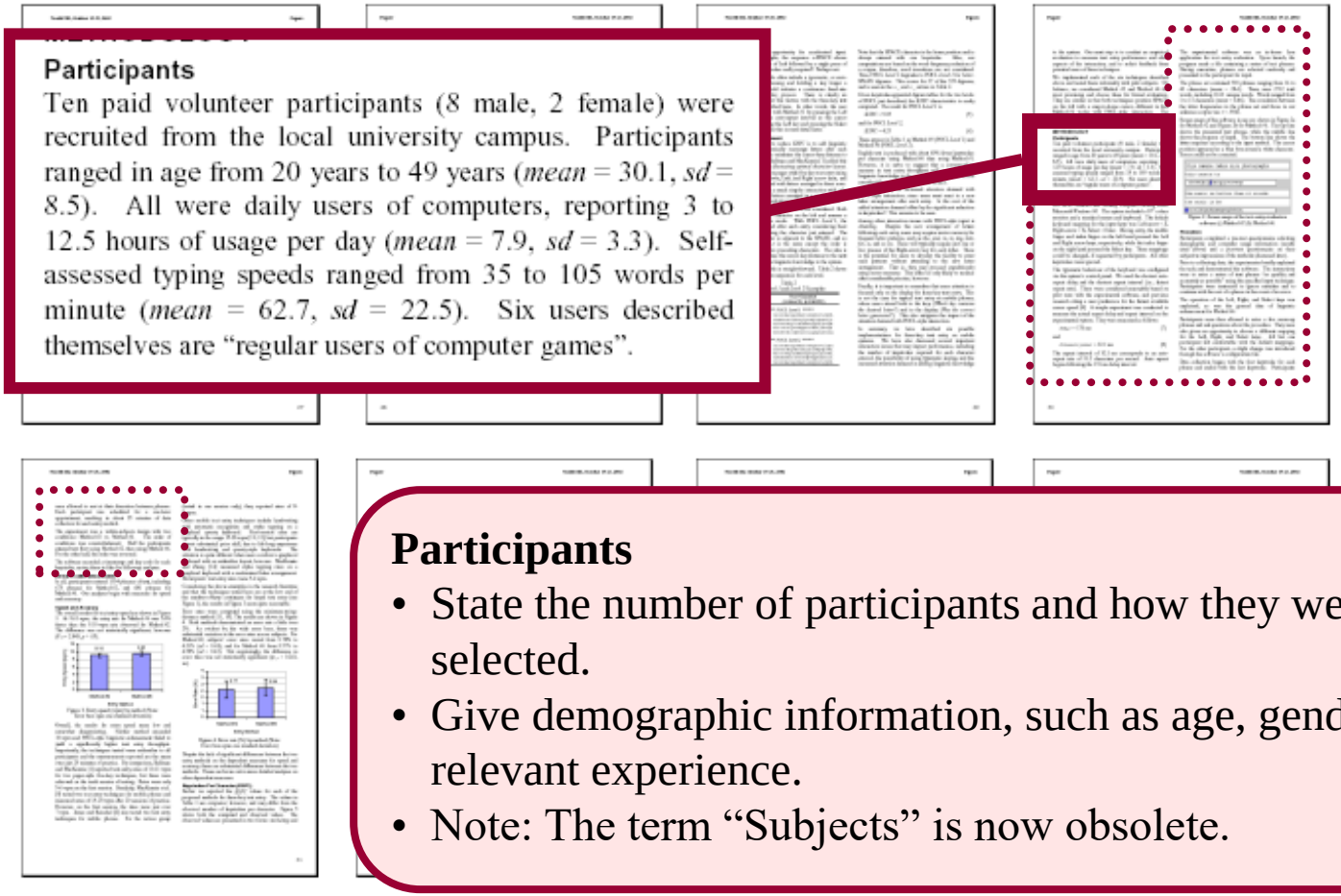
Method



Method

- Tell the reader **what you did** and **how you did it**.
- The research must be reproducible!
- Use the following subsections…

Method - Participants



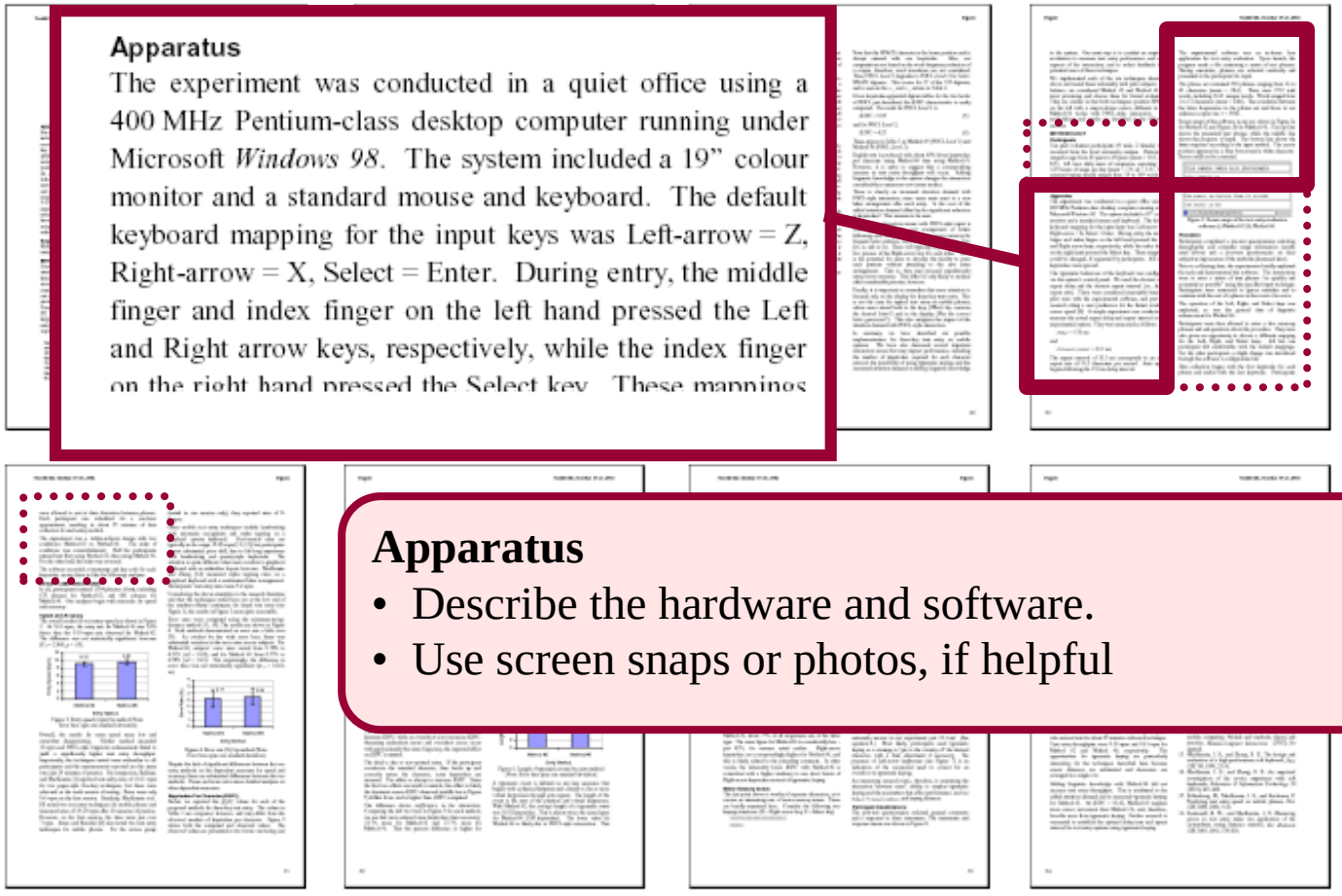
Participants

Ten paid volunteer participants (8 male, 2 female) were recruited from the local university campus. Participants ranged in age from 20 years to 49 years (*mean* = 30.1, *sd* = 8.5). All were daily users of computers, reporting 3 to 12.5 hours of usage per day (*mean* = 7.9, *sd* = 3.3). Self-assessed typing speeds ranged from 35 to 105 words per minute (*mean* = 62.7, *sd* = 22.5). Six users described themselves are "regular users of computer games".

Participants

- State the number of participants and how they were selected.
- Give demographic information, such as age, gender, relevant experience.
- Note: The term "Subjects" is now obsolete.

Method - Apparatus



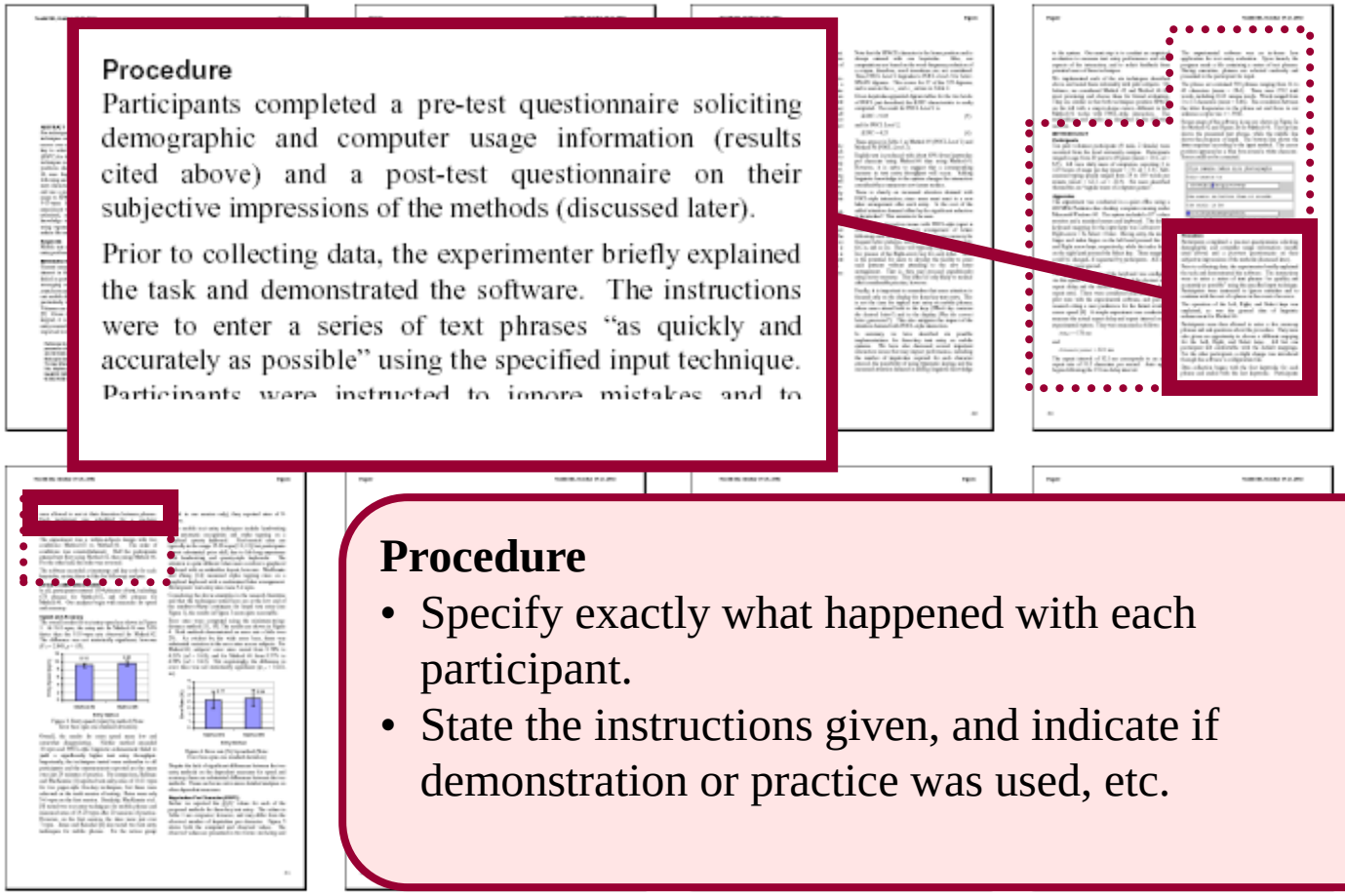
Apparatus

The experiment was conducted in a quiet office using a 400 MHz Pentium-class desktop computer running under Microsoft *Windows 98*. The system included a 19" colour monitor and a standard mouse and keyboard. The default keyboard mapping for the input keys was Left-arrow = Z, Right-arrow = X, Select = Enter. During entry, the middle finger and index finger on the left hand pressed the Left and Right arrow keys, respectively, while the index finger on the right hand pressed the Select key. These mappings

Apparatus

- Describe the hardware and software.
- Use screen snaps or photos, if helpful

Method - Procedure



Procedure

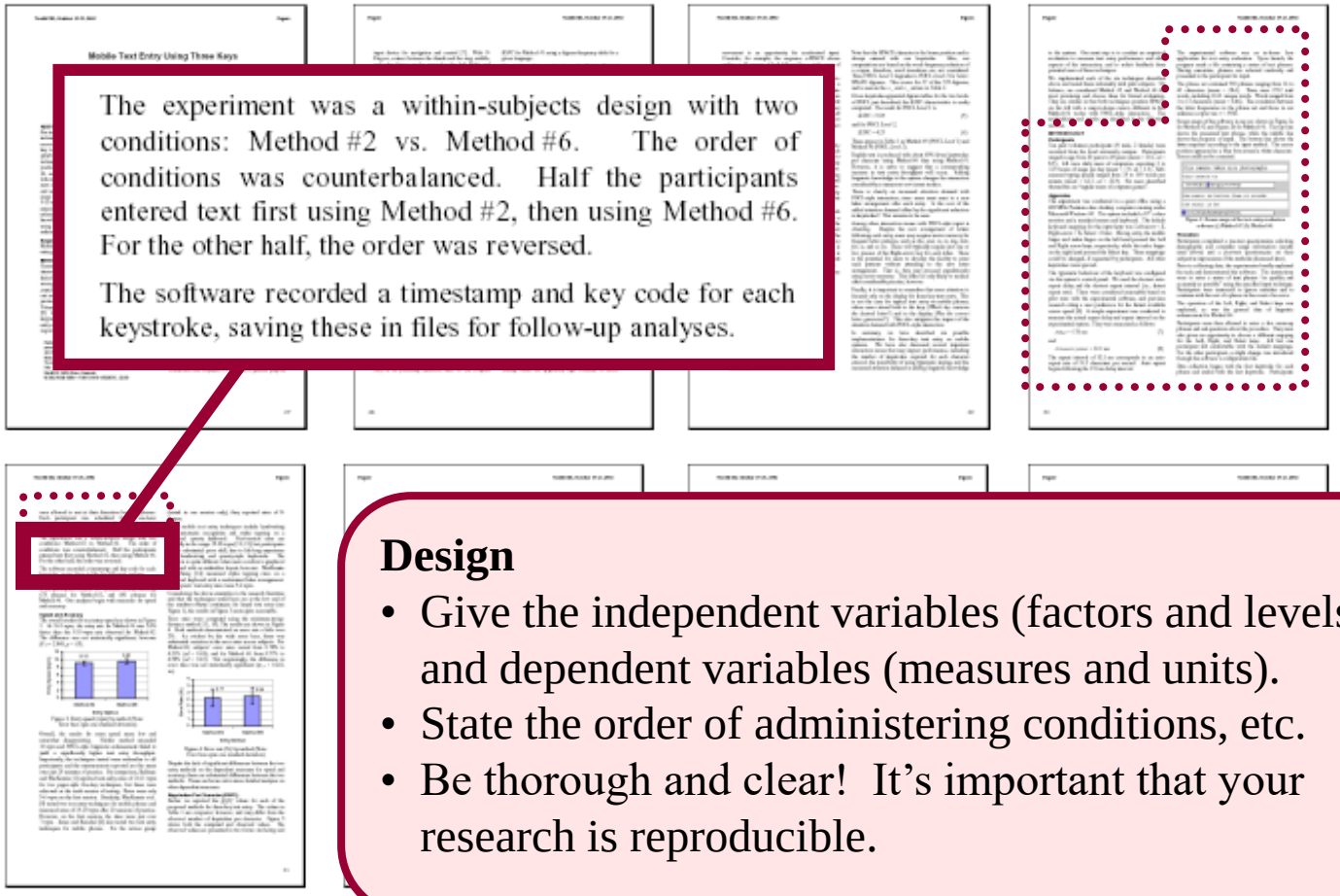
Participants completed a pre-test questionnaire soliciting demographic and computer usage information (results cited above) and a post-test questionnaire on their subjective impressions of the methods (discussed later).

Prior to collecting data, the experimenter briefly explained the task and demonstrated the software. The instructions were to enter a series of text phrases "as quickly and accurately as possible" using the specified input technique. Participants were instructed to ignore mistakes and to

Procedure

- Specify exactly what happened with each participant.
- State the instructions given, and indicate if demonstration or practice was used, etc.

Method - Design



The experiment was a within-subjects design with two conditions: Method #2 vs. Method #6. The order of conditions was counterbalanced. Half the participants entered text first using Method #2, then using Method #6. For the other half, the order was reversed.

The software recorded a timestamp and key code for each keystroke, saving these in files for follow-up analyses.

Design

- Give the independent variables (factors and levels) and dependent variables (measures and units).
- State the order of administering conditions, etc.
- Be thorough and clear! It's important that your research is reproducible.

Results and Discussion

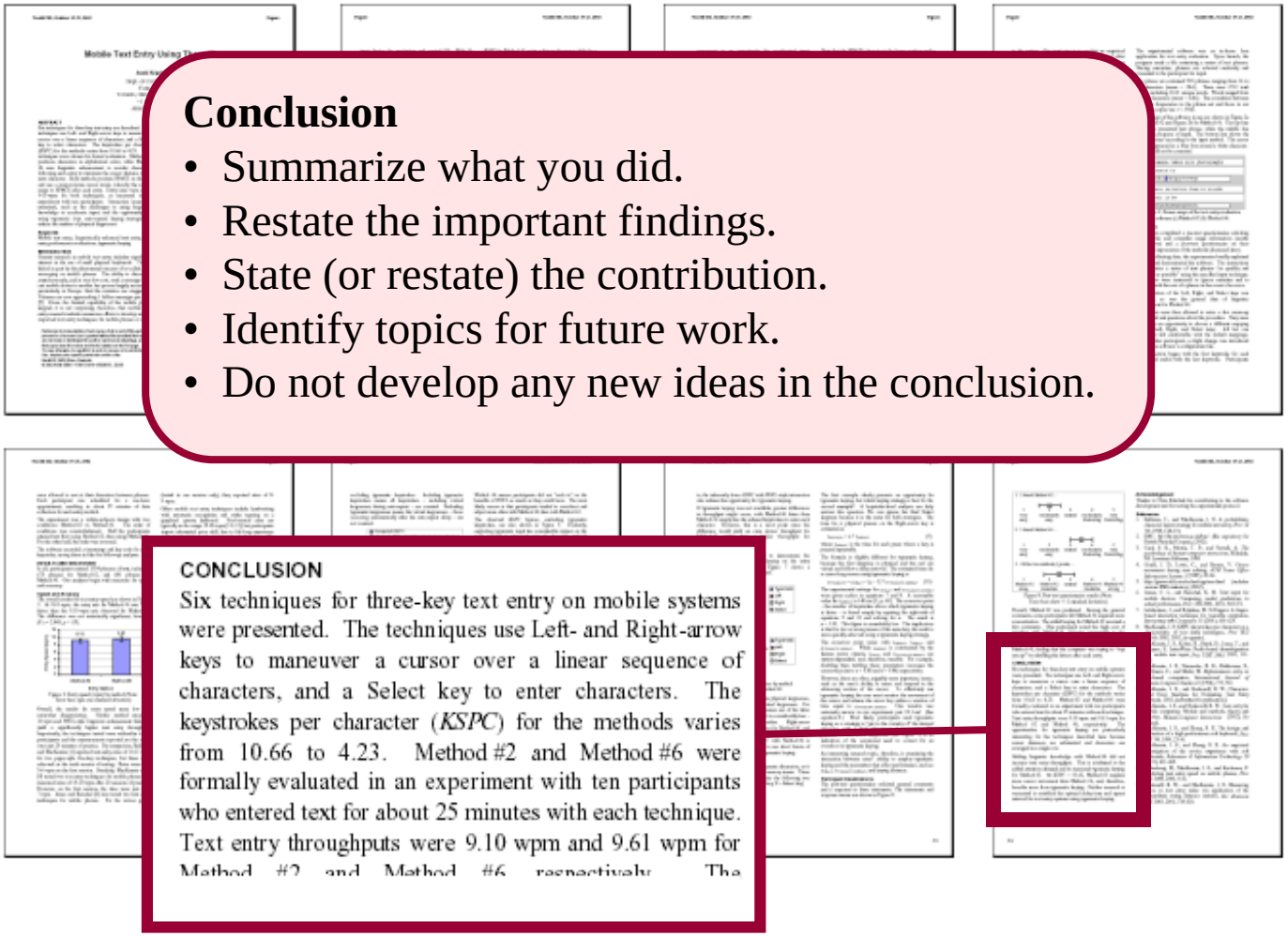
Results and Discussion

- Use subsections as appropriate
- If there were outliers or problems in the data collection, state this up-front.
- Organize results by the dependent measures, moving from overall means to finer details across conditions.
- Use statistical tests, charts, tables, as appropriate

Results and Discussion (2)

- Don’t overdo it! Giving too many charts or too much data means you can’t distinguish what is important from what is not important.
- Discuss the results. State what is interesting
- Explain the differences across conditions.
- Compare with results from other studies.
- Provide additional analysis, as appropriate, such as fine grain analyses on types of errors or linear regression or correlation analyses for models of interaction (such as Fitts’ law).

Conclusion



Conclusion

- Summarize what you did.
- Restate the important findings.
- State (or restate) the contribution.
- Identify topics for future work.
- Do not develop any new ideas in the conclusion.

CONCLUSION

Six techniques for three-key text entry on mobile systems were presented. The techniques use Left- and Right-arrow keys to maneuver a cursor over a linear sequence of characters, and a Select key to enter characters. The keystrokes per character (*KSPC*) for the methods varies from 10.66 to 4.23. Method #2 and Method #6 were formally evaluated in an experiment with ten participants who entered text for about 25 minutes with each technique. Text entry throughputs were 9.10 wpm and 9.61 wpm for Method #2 and Method #6 respectively. The

Reference and source

- Conceptual Econometrics Using R (ISSN Book 41) 1st Edition, by Hrishikesh D. Vinod (Editor)
- Principles of Macroeconometric Modeling (Volume 36) (Advanced Textbooks in Economics, Volume 36) by L.R. Klein, W. Welfe, et al. | Oct 5, 1999
- Macroeconomic Modeling and Macroeconometric Simulation: Illustrated with a developing economy Model (Macroeconometric model Book 1) Book 1 of 1: Macroeconometric model | by Kannapiran Arjunan | Jun 9, 2020
- Global and National Macroeconometric Modelling: A Long-Run Structural Approach by Anthony Garratt, Kevin Lee, et al. | May 4, 2012
- Simulation of a macroeconometric model with multiple time series considerations (Wayne economic papers) by Rosemary Rossiter | Jan 1, 1982