Leibniz
Universität
Hannover

Mensch-Computer-Interaktion 2

Experiments

Human-Computer
Interaction Group

Prof. Dr. Michael Rohs
michael.rohs@hci.uni-hannover.de

Lectures

Session	Date	Topic	
1	6.4.	Introduction	
2	13.4.	Interaction elements	GUI toolkits, interaction techniques
3	20.4.	Event handling	
4	27.4.	Scene graphs	
5	4.5.	Interaction techniques	
	11.5.	no class (CHI)	
	18.5.	no class (spring break)	
6	25.5.	Experiments	design and analysis of experiments
7	1.6.	Data Analysis	
8	8.6.	Data Analysis	
9	15.6.	Visualization	
10	22.6.	Visualization	current topics beyond-desktop UIs
11	29.6.	Modeling interaction	
12	6.7.	Computer vision for interaction	
13	13.7.	Computer vision for interaction	

Klausur:
28.7.2016
8-11 Uhr
HG E214

Research Methods

- Observational method
- Experimental method
- Correlational method

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Observational Method

- Example methods
  - Interviews, field investigations, contextual inquiries, case studies, field studies, focus groups, think aloud protocols, story telling, walkthroughs, cultural probes, etc.
- Focus on qualitative assessments (cf. quantitative)
  - “Why” or “how” of the interaction
  - Attitude, emotion, strategy, etc.
- Relevance vs. precision
  - High in relevance (behaviors studied in a natural setting)
  - Low in precision (lacks control available in a laboratory)
- Goal: Discover and explain reasons underlying human behavior

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Experimental Method

- Controlled experiments conducted in lab setting
- Relevance vs. precision
  - Low in relevance (artificial environment)
  - High in precision (extraneous behaviours easy to control)
- At least two variables
  - Manipulated variable (aka independent variable)
  - Response variable (aka dependent variable)
- Cause-and-effect conclusions possible
  - Changes in the manipulated variable caused changes in the response variable

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Correlational Method

- Look for relationships between variables
- Observations made, data collected
  - Example: Are user's privacy settings in social networking related to their age, gender, level of education, employment status, income, etc.
- Non-experimental
  - Interviews, on-line surveys, questionnaires, etc.
- Balance between relevance and precision
  - Some quantification, observations not in lab
- Cause-and-effect conclusions not possible
  - Example: Shoe size and reading ability

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

DESIGNING EXPERIMENTS

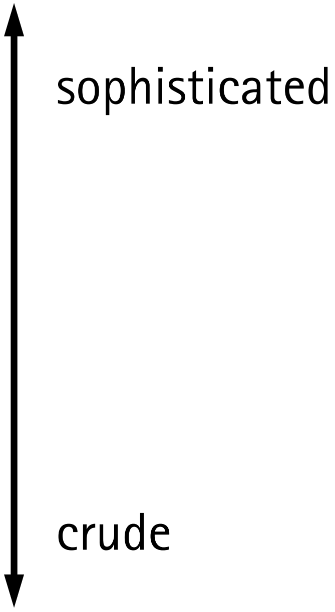
Observe and Measure

- Foundation of empirical research
- Observation is the starting point; observations are made
  - By the apparatus
  - By a human observer
- Manual observation
  - Log sheet, notebooks
  - Screen capture, photographs, videos, etc.
- Measurement
  - With measurement, anecdotes (April showers bring May flowers) turn to empirical evidence
  - “When you cannot measure, your knowledge is of a meagre and unsatisfactory kind” (Kelvin)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Scales of Measurement

- Ratio
- Interval
- Ordinal
- Nominal



sophisticated

crude

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Nominal Data

- Nominal data (aka categorical data) are arbitrary codes assigned to attributes
  - 1 = male, 2 = female
  - 1 = mouse, 2 = touchpad, 3 = pointing stick
- The code needn't be a number
  - M = male, F = female
- 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 handers or right handers have more difficulty with…"
  - Note: The count itself is a ratio-scale measurement

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Nominal Data – HCI Example

- Task: 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%**		

males on phone: 12.5 %
females on phone: 13.7 %

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Ordinal Data

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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Ordinal Data – HCI Example

- Example: 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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Interval Data

- Equal distances between boundaries of intervals of same size
- But no absolute zero
- Classic example: Temperature in °C
- Statistical mean possible
  - Example: Mean midday temperature during July
- Ratios not possible
  - Cannot say 10 °C is twice as warm as 5 °C
  - 10 °C is minus ten times as warm as -1 °C (?)
  - 10 °C is infinitely warmer than 0 °C (?)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Interval Data – HCI Example

- Questionnaires often solicit a level of agreement to a statement
- Responses on a Likert scale
- Likert scale characteristics
  - Level of agreement to a statement
  - Responses are symmetric about a neutral middle value
  - Gradations between responses are equal (more-or-less)
- Assuming "equal gradations", the statistical mean is valid
  - (and related statistical tests are possible)
- Likert scale example

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Interval Data – HCI Example (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 read a text message 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

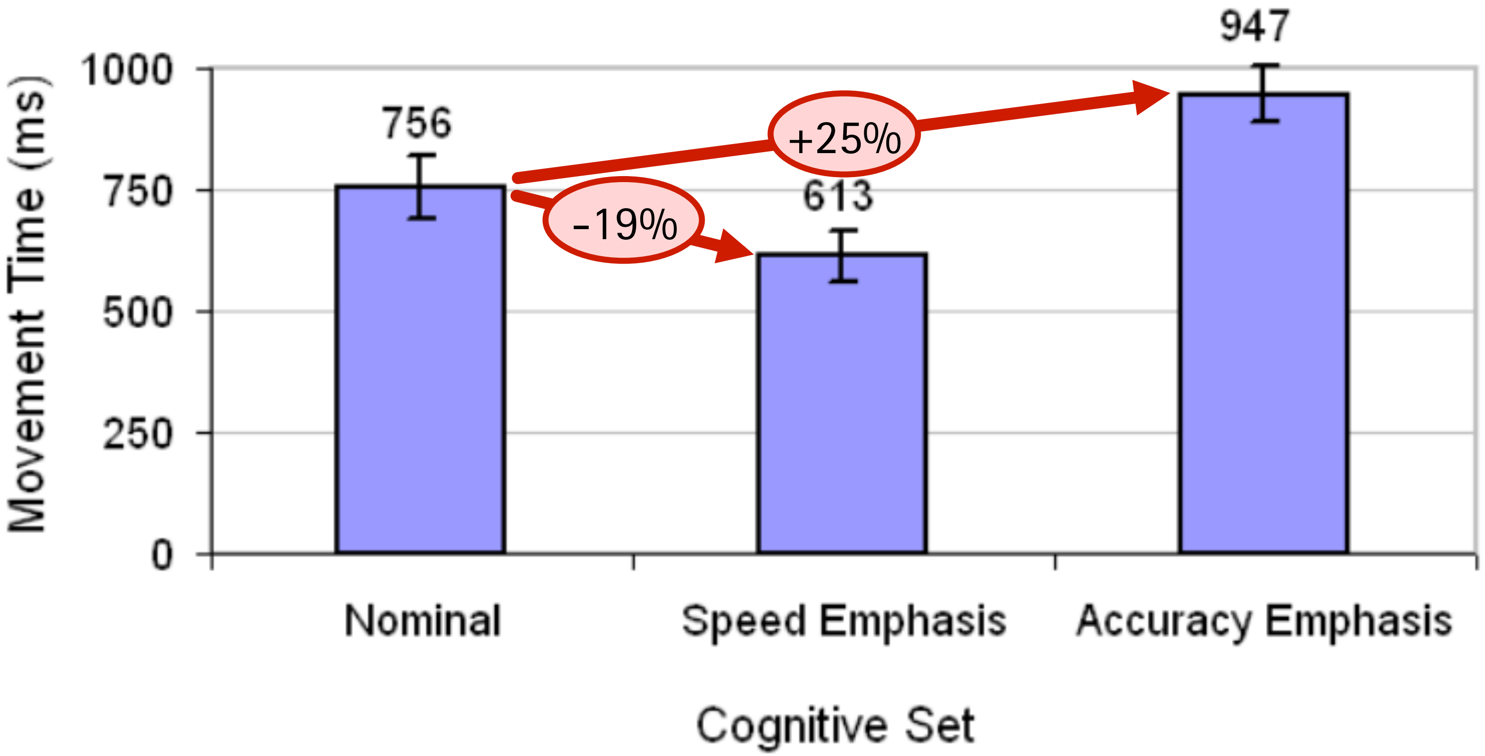
MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Ratio Data

- Preferred scale of measurement
- Absolute zero, therefore many calculations possible
- Ratios enable summaries and comparisons
- A “count” is a ratio-scale measurement
  - Example: Error rate (count of wrong trials relative to count of all trials)
- Enhance counts by adding further ratios where possible
  - Normalization facilitates comparisons
  - Example: A 10-word phrase was entered in 30 seconds
    - Bad: t = 30 seconds
    - Good: Entry rate = 10 / 0.5 = 20 wpm

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Ratio Data – HCI Example[1]



Ratio: 947 / 756 = 1.253 = 125%

Ratio: 613 / 756 = 0.811, 1 - 0.811 = 0.189 = 19%

[1] MacKenzie, Isokoski. Fitts' throughput and the speed-accuracy tradeoff. Proc. CHI 2008.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Experiments in HCI

- Learning to conduct and design an experiment is a skill required of all researchers in HCI

- Experiment design is the process of deciding
  what variables to use,
  what tasks and procedures to use,
  how many participants to have, and
  how to solicit them, and so on

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Getting Started With Experiment Design

- Creative work (ideas, prototypes) → Experimental research
  - Transitioning from the creative work to experiments is a challenge
- Begin with…

> What are the experimental variables?

- Remember research questions

> Can a task be performed more quickly with my new interface than with an existing interface?

- Properly formed research questions inherently identify experimental variables (above?)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Testable Research Questions

- Questions may be relevant but not testable
  - Is it better than current practice?
  - Which design alternative is best?
  - What are the weaknesses?
  - How much practice is required?
- Try to re-cast as testable questions
  - (even though the new question may appear less important)
- Scenario: You have invented a new text entry technique for touchscreens. You think it is better than the Qwerty soft keyboard (QSK). You decide to evaluate your invention.
  - What are your research questions?

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Testable Research Questions for new Text Entry Technique

- Very weak

  Is the new technique any good?
- Weak

  Is the new technique better than QSK?
- Better

  Is the new technique faster than QSK?
- Better still

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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Internal Validity

- Definition
  - The extent to which the effects observed are due to the test conditions (e.g., multitap vs. new method)
- Statistically, this means…
  - Differences are due to inherent properties of the test conditions
  - Variations are due to participant differences (“pre-dispositions”)
  - Other potential sources of variation are controlled or exist equally or randomly across the test conditions
- Threats to internal validity
  - One group tested early, the other late in the day
  - One group has lots of experience with the task, the other has not
  - etc.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Test Environment Example

- Scenario: Compare two input devices for pointing at a projection screen
- External validity is improved if the test environment is similar to expected usage environment
- A realistic test environment should probably
  - Use a large display or projection screen (not a desktop monitor)
  - Have participants stand (rather than sit)
  - Include an audience!
- But... is internal validity compromised?

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

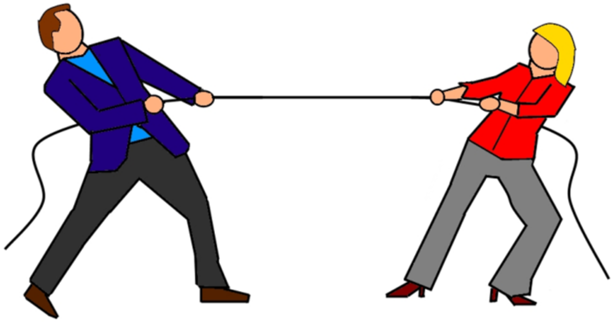
Experimental Procedure Example

- Scenario: 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 personalized paragraphs of text (e.g., a paragraph about a favorite movie)
  - Edit and correct mistakes as they normally would
- But… is internal validity compromised?

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

The Tradeoff

Internal validity



External validity

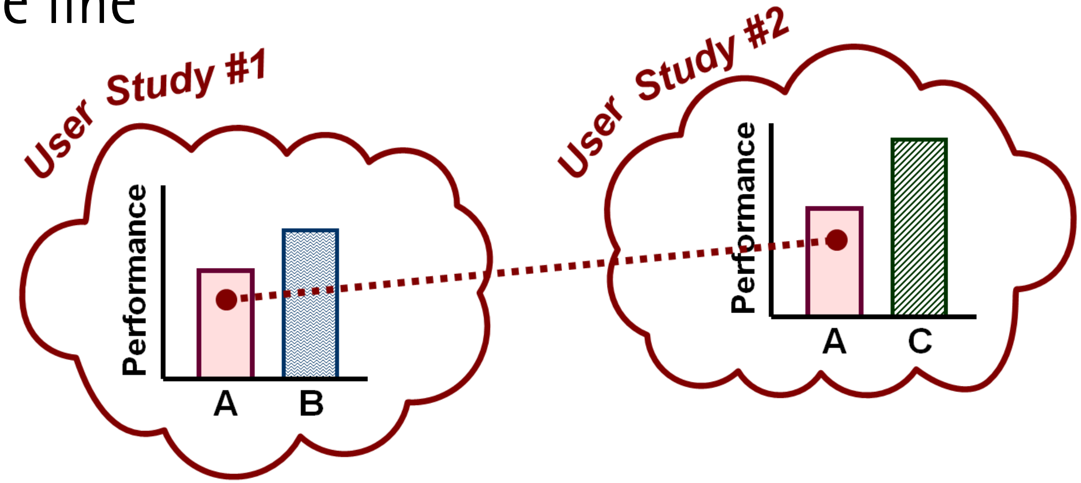
- There is tension between internal and external validity
- The more the test environment and experimental procedures are “relaxed” (to mimic real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as pondering, distractions, fiddling, or secondary tasks

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Comparative Evaluations

- Comparative evaluation is preferable to a one-of
  - E.g., input method A vs. input method B
  - More insightful results obtained
- Comparative evaluations have at least one independent variable with at least two levels
- If one condition is a base line comparisons possible between studies

  (assuming similar methodology)



MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Independent Variable

- An independent variable (IV) is a circumstance or characteristic that is manipulated in an experiment to elicit a change in a human response while interacting with a computer
- “Independent” because it is independent of participant behavior
  - There is nothing a participant can do to influence an independent variable
- Examples
  - Interface, device, feedback mode, button layout, visual layout, age, gender, background noise, expertise, etc.
- The terms independent variable and factor are synonymous

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Test Conditions

- An independent variable (IV) must have at least two levels
- The levels, values, or settings for an IV are the test conditions
- Name both the factor (IV) and its levels (test conditions):

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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Human Characteristics

- Human characteristics are naturally occurring attributes
- Examples
  - Gender, age, height, weight, handedness, grip strength, finger width, visual acuity, personality trait, first language, etc.
- They are legitimate independent variables, but they cannot be "manipulated" in the usual sense
- Causal relationships are difficult to obtain due to unavoidable confounding variables

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Dependent Variable

- A dependent variable is a measured human behavior
  - Potentially related to an independent variable
- "Dependent" because it depends on what the participant does
- "Dependent" because it might depend on the independent variable
- Examples
  - Task completion time, speed, accuracy, error rate, throughput, target re-entries, task retries, presses of backspace, etc.
- Dependent variables must be clearly defined
  - Research must be reproducible!

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Unique Dependent Variables

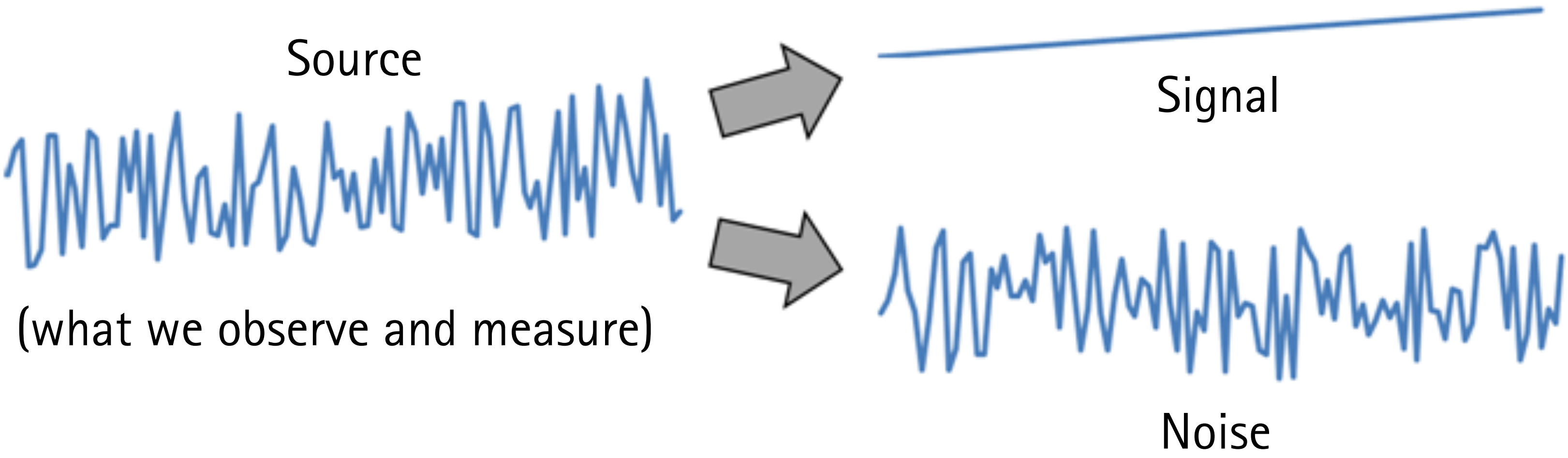
- Any observable, measurable behaviour is a legitimate DV
  (provided it has the potential to reveal differences among the test conditions)
- So, feel free to "roll your own"
- Example: Negative facial expressions[1]
  - Application: User difficulty with mobile games
  - Events logged included frowns, head shaking
  - Clearly defined → reproducible

[1] Duh, Chen, Tan. Playing different games on different phones: An empirical study on mobile gaming. Proc. MobileHCI 2008.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Signal and Noise Metaphor

- Signal and noise metaphor for experiment design:



- Signal → a (dependent) variable of interest
- Noise → everything else (random influences)
- Experiment design seeks to enhance the signal, while minimizing the noise

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

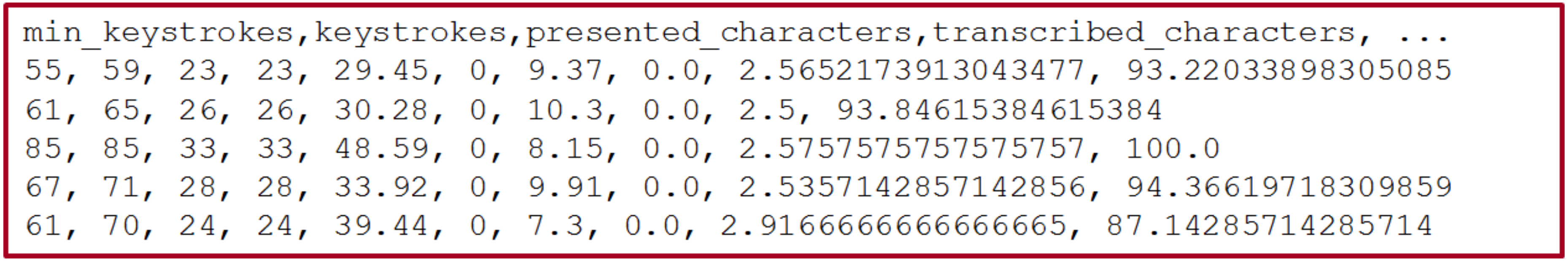
Data Collection

- Data for dependent variables must be collected somehow
- Ideally, experiment software logs timestamps, key presses, button clicks, etc.
- Planning and pilot testing experiment software is important
- Ensure conditions are identified in the filenames and in the data columns

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Data Collection Example

TextInputHuffman-P01-D99-B06-S01.sd2

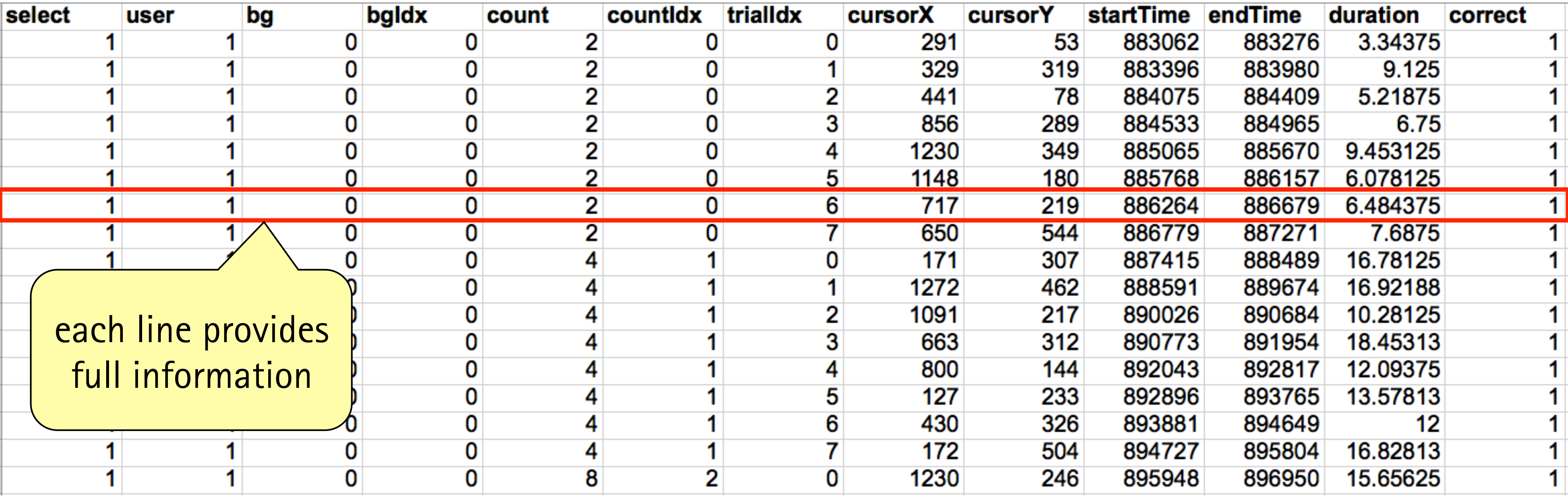


```
min_keystrokes,keystrokes,presented_characters,transcribed_characters, ...
55, 59, 23, 23, 29.45, 0, 9.37, 0.0, 2.5652173913043477, 93.22033898305085
61, 65, 26, 26, 30.28, 0, 10.3, 0.0, 2.5, 93.84615384615384
85, 85, 33, 33, 48.59, 0, 8.15, 0.0, 2.5757575757575757, 100.0
67, 71, 28, 28, 33.92, 0, 9.91, 0.0, 2.5357142857142856, 94.36619718309859
61, 70, 24, 24, 39.44, 0, 7.3, 0.0, 2.9166666666666665, 87.14285714285714
```

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Data Collection Example

- Tabular form, redundancy is not an issue
- Ease of analysis is an issue
- Each line has full information
  - user, trial, start/end time of trial, state of IVs, measured DVs



select	user	bg	bgIdx	count	countIdx	trialIdx	cursorX	cursorY	startTime	endTime	duration	correct
1	1	0	0	2	0	0	291	53	883062	883276	3.34375	1
1	1	0	0	2	0	1	329	319	883396	883980	9.125	1
1	1	0	0	2	0	2	441	78	884075	884409	5.21875	1
1	1	0	0	2	0	3	856	289	884533	884965	6.75	1
1	1	0	0	2	0	4	1230	349	885065	885670	9.453125	1
1	1	0	0	2	0	5	1148	180	885768	886157	6.078125	1
1	1	0	0	2	0	6	717	219	886264	886679	6.484375	1
1	1	0	0	2	0	7	650	544	886779	887271	7.6875	1
1	1	0	0	4	1	0	171	307	887415	888489	16.78125	1
[illegible]	[illegible]	0	0	4	1	1	1272	462	888591	889674	16.92188	1
[illegible]	[illegible]	0	0	4	1	2	1091	217	890026	890684	10.28125	1
[illegible]	[illegible]	0	0	4	1	3	663	312	890773	891954	18.45313	1
[illegible]	[illegible]	0	0	4	1	4	800	144	892043	892817	12.09375	1
[illegible]	[illegible]	0	0	4	1	5	127	233	892896	893765	13.57813	1
[illegible]	[illegible]	0	0	4	1	6	430	326	893881	894649	12	1
1	1	0	0	4	1	7	172	504	894727	895804	16.82813	1
1	1	0	0	8	2	0	1230	246	895948	896950	15.65625	1

each line provides full information

Control Variable

- Control variable: Circumstance (not under investigation) that is kept constant while testing the effect of an independent variable
- More control means the experiment is less generalizable
  - I.e., less applicable to other people and other situations
- Research question: Is there an effect of font color or background color on reading comprehension?
  - Independent variables: font color, background color
  - Dependent variable: comprehension test scores
  - Control variables
    - Font size (e.g., 12 point)
    - Font family (e.g., Times)
    - Ambient lighting (e.g., fluorescent, fixed intensity)
    - etc.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Random Variable

- 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!)
- Research question: Does user stance affect 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.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Control vs. Random Variables

There is a trade-off which can be examined in terms of internal validity and external validity

Variable	Advantage	Disadvantage
Random	Improves external validity by using a variety of situations and people	Compromises internal validity by introducing additional variability
Control	Improves internal validity since variability due to a controlled circumstance is eliminated	Compromises external validity by limiting responses to specific situations and people

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Confounding Variable

- Confounding variable: Circumstance that varies systematically with an independent variable
  - Should be considered, lest the results are misleading
- Research question: In an eye tracking application, is there an effect of "camera distance" on task completion time?
  - Independent variable: Camera distance (near, far)
    - Near camera (A): inexpensive camera mounted on eye glasses
    - Far camera (B): expensive camera mounted above large display
  - Dependent variable: task completion time
  - Camera is confounding variable: camera A near, camera B far
  - Effects due to camera distance or due to camera quality?

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Experiment Task

- Recall the definition of an independent variable
  - A circumstance or characteristic that is manipulated in an experiment to elicit a change in a human response while interacting with a computer
- The experiment task must potentially "elicit a change"
- Qualities of a good task: represent, discriminate
  - Represent activities people do with the interface
    - Improves external validity (but may compromise internal validity)
  - Discriminate among the test conditions
    - Increases likelihood of a statistically significant outcome (i.e., the sought-after "change" occurs)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Task Examples

- Usually the task is self-evident
    - Usually follows directly from the research idea
- Research idea → A new graphical method for entering equations in a spreadsheet
    - Experiment task → Insert an equation using (a) the graphical method and (b) the conventional method
- Research idea → An auditory feedback technique for programming a GPS device
    - Experiment task → Program a destination location using (a) the auditory feedback method and (b) the conventional method

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Knowledge-Based Tasks

- Most experiment tasks are performance-based or skill-based
  - Example: Inserting an equation, programming a destination location
- Sometimes the task is knowledge-based
  - Example: “Use an Internet search interface to find the birth date of Albert Einstein.”
  - Participants become “contaminated” after the first run of task, since they have acquired the knowledge
- Experimentally, knowledge-based tasks pose problems
- A creative approach is needed
  - Example: Slightly change the task; “...of William Shakespeare”

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Procedure

- The procedure encompasses everything that occurs with participants
- The procedure includes the experiment task and everything else before / after / between
  - Arriving, welcoming
  - Signing a consent form
  - Instructions given to participants about the experiment task
  - Demonstration trials, practice trials
  - Pauses
  - Administering of a questionnaire or an interview

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Instructions

- Very important (prepare in advance, written)
- Often the goal in the experiment task is “to proceed as quickly and accurately as possible but at a pace that is comfortable”
- Give the (exact!) same instructions to all participants
- If a participant asks for clarification, do not change the instructions in a way that may cause the participant to behave differently from the other participants

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Participants

- Researchers want experimental results to apply to people not actually tested – a population
- Population examples
  - Computer-literate adults, teenagers, children, people with certain disabilities, left-handed people, engineers, musicians, etc.
- For results to apply generally to a population, the participants used in the experiment must be...
  - Members of the desired population
  - Selected at random from the population
- True random sampling is rarely done
- Some form of convenience sampling is typical

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

How Many Participants?

- Too few → Experimental effects fail to achieve statistical significance
- Too many → Statistical significance for effects of no practical value
- The correct number… (drum roll please)
  - Use the same number of participants as used in similar research[1]
  - Difficult to derive analytically or estimate
  - Depends on effect size

[1] Martin. Doing psychology experiments (6th ed.). Pacific Grove, CA. Belmont, CA: Wadsworth, 2004.

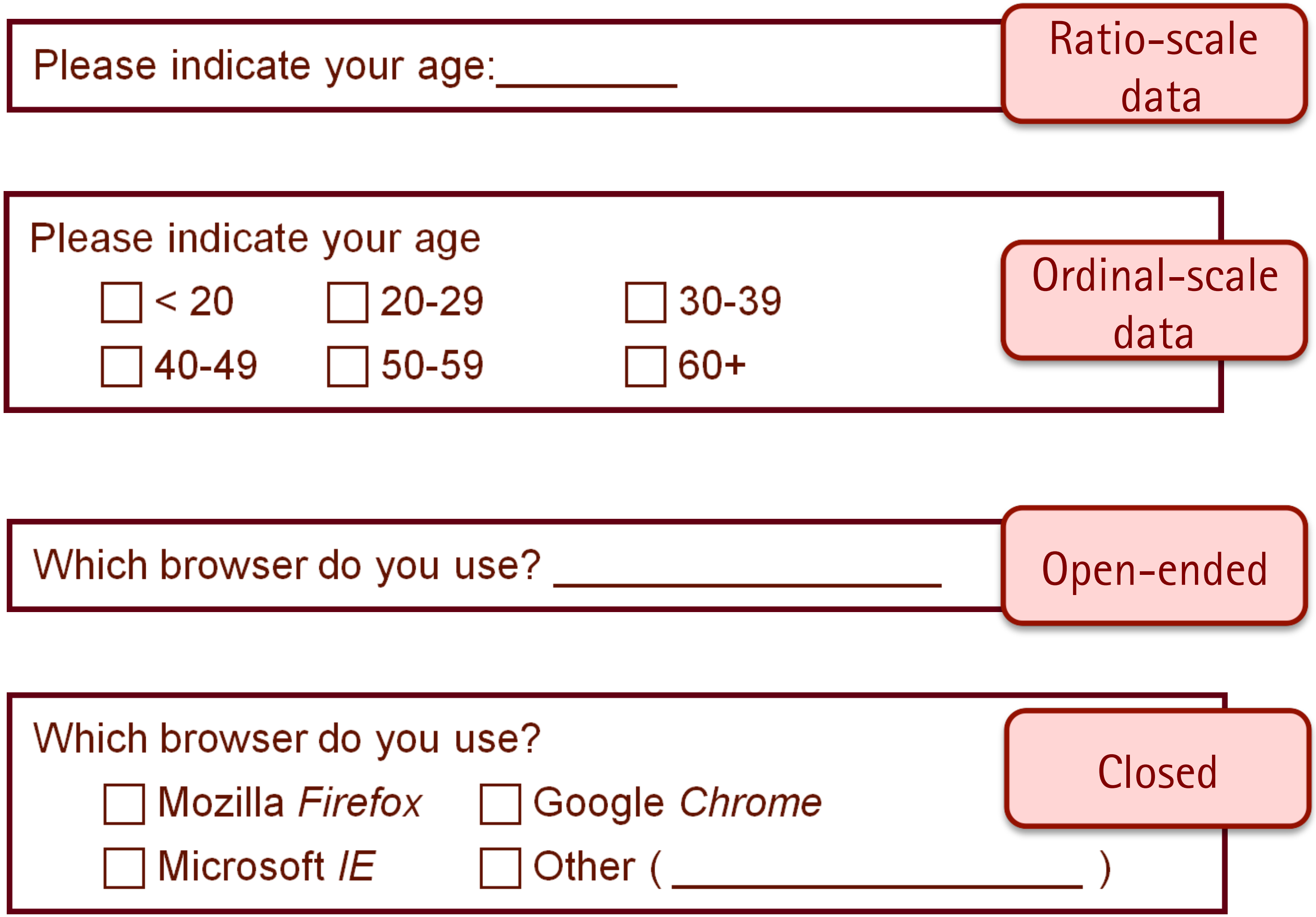
MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Questionnaires

- Questionnaires are used in most HCI experiments
- Two purposes: Demographics, feedback
  - Collect information about the participants
    - Demographics (gender, age, first language, handedness, visual acuity, etc.)
    - Prior experience with interfaces or interaction techniques related to the research
  - Solicit feedback, comments, impressions, suggestions, etc., about participants' use of the experimental apparatus
- Questionnaires, as an adjunct to experimental research, are usually brief

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Information Questions



Please indicate your age:________

Ratio-scale data

Please indicate your age

□ < 20 □ 20-29 □ 30-39

□ 40-49 □ 50-59 □ 60+

Ordinal-scale data

Which browser do you use? ________________

Open-ended

Which browser do you use?

□ Mozilla *Firefox* □ Google *Chrome*

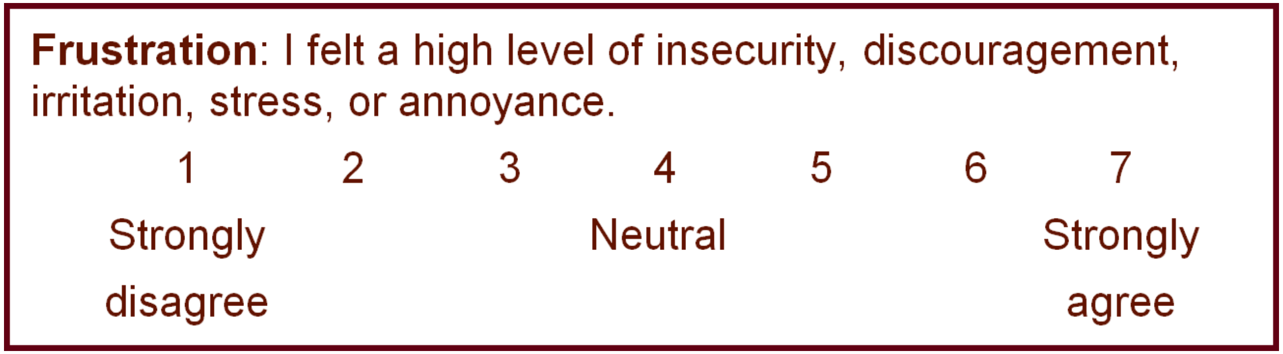
□ Microsoft *IE* □ Other (________________)

Closed

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Participant Feedback

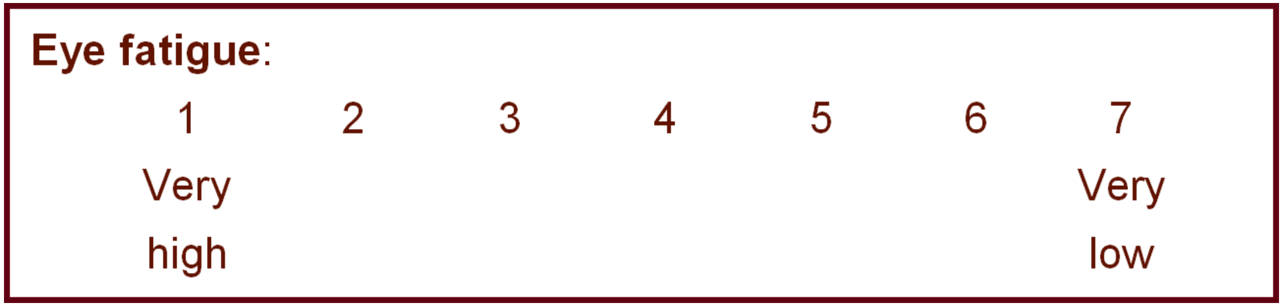
- Using NASA Task Load Index (TLX):



Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

1	2	3	4	5	6	7
Strongly disagree			Neutral			Strongly agree

- ISO 9241-9:



Eye fatigue:

1	2	3	4	5	6	7
Very high						Very low

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Within-Subjects, Between-Subjects

- Two ways to assign conditions to participants
  - Within-subjects → Each participant is tested on each condition
  - Between-subjects → Each participant is tested on one condition only
  - Example: An IV with three test conditions (A, B, C):

Within-subjects

Participant	Test Condition		
1	A	B	C
2	A	B	C

Between-subjects

Participant	Test Condition
1	A
2	A
3	B
4	B
5	C
6	C

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Within-Subjects, Between-Subjects (2)

- Within-subjects advantages
  - Fewer participants (easier to recruit, schedule, etc.)
  - Less “variation due to participants”
  - No need to balance groups (because there is only one group)
- Within-subjects disadvantage
  - Order effects (i.e., interference between conditions)
- Between-subjects advantage
  - No order effects (i.e., no interference between conditions)
- Between-subjects disadvantage
  - More participants (harder to recruit, schedule, etc.)
  - More “variation due to participants”
  - Need to balance groups (to ensure they are more or less the same)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Within-Subjects, Between-Subjects (3)

- Sometimes…
  - A factor must be assigned within-subjects
    - Examples: Block, session (if learning is the IV)
  - A factor must be assigned between-subjects
    - Examples: gender, handedness
  - There is a choice
    - In this case, within-subjects is often better
- With two factors, there are three possibilities
    - Both factors within-subjects
    - Both factors between-subjects
    - One factor within-subjects + one factor between-subjects (this is a mixed design)

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Order Effects, Counterbalancing

- Only relevant for within-subjects factors
- The issue: Order effects
  - Aka: Learning effects, practice effects, fatigue effects, sequence effects
- Order effects offset by counterbalancing
  - Participants divided into groups
  - Test conditions are administered in a different order to each group
  - Order of administering test conditions uses all possible permutations (n! for n conditions) or a subset of the possible permutations (Latin square)
- Latin square
  - Each condition occurs precisely once in each row and column

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Latin Squares

2 x 2

A	B
B	A

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

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

5 x 5

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

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Balanced Latin Square

- With a balanced Latin square, each condition precedes and follows each other condition an equal number of times
  - Only possible for even-orders
  - Top row pattern: A, B, n, C, n – 1, D, n – 2, ...
  - Columns ordered (A, B, ..., n), with wrap around

4 x 4

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

6 x 6

A	B	F	C	E	D
B	C	A	D	F	E
C	D	B	E	A	F
D	E	C	F	B	A
E	F	D	A	C	B
F	A	E	B	D	C

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Example

- An experimenter seeks to determine if three editing methods (A, B, C) differ in the amount of time to do an editing task:

> Replace one 5-letter word with another, starting one line away.

  - Method A: Arrow keys, backspace, type
  - Method B: Search and replace dialog
  - Method C: Point and double click, type

one IV with 3 levels

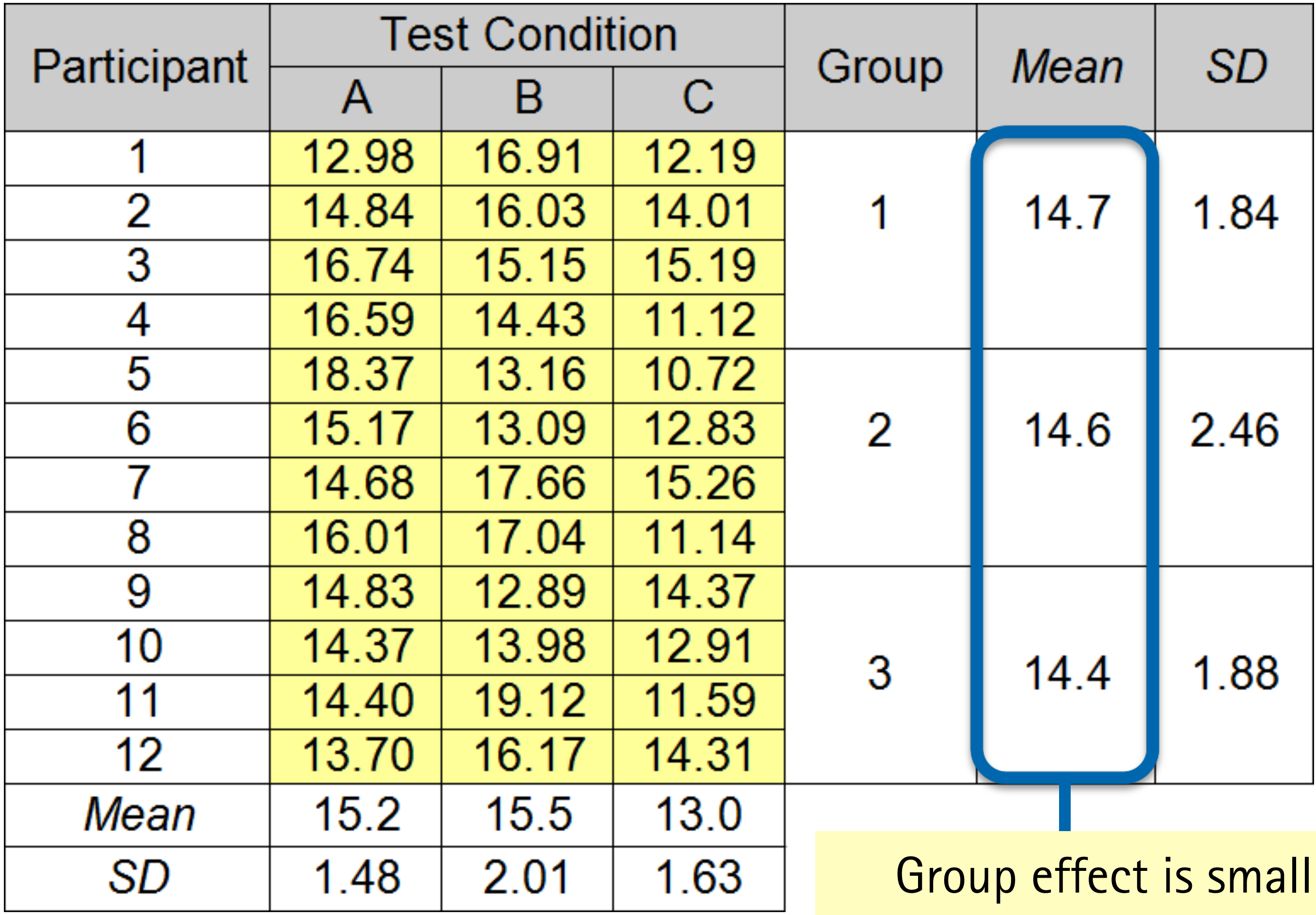
- Conditions are assigned within-subjects
- Twelve participants are recruited and divided into three groups

  4 participants/group

- Methods administered using a 3 × 3 Latin square

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Results – Data



Participant	Test Condition			Group	*Mean*	*SD*
	A	B	C			
1	12.98	16.91	12.19	1	14.7	1.84
2	14.84	16.03	14.01			
3	16.74	15.15	15.19			
4	16.59	14.43	11.12			
5	18.37	13.16	10.72	2	14.6	2.46
6	15.17	13.09	12.83			
7	14.68	17.66	15.26			
8	16.01	17.04	11.14			
9	14.83	12.89	14.37	3	14.4	1.88
10	14.37	13.98	12.91			
11	14.40	19.12	11.59			
12	13.70	16.17	14.31			
Mean	15.2	15.5	13.0			
SD	1.48	2.01	1.63			

order: ABC

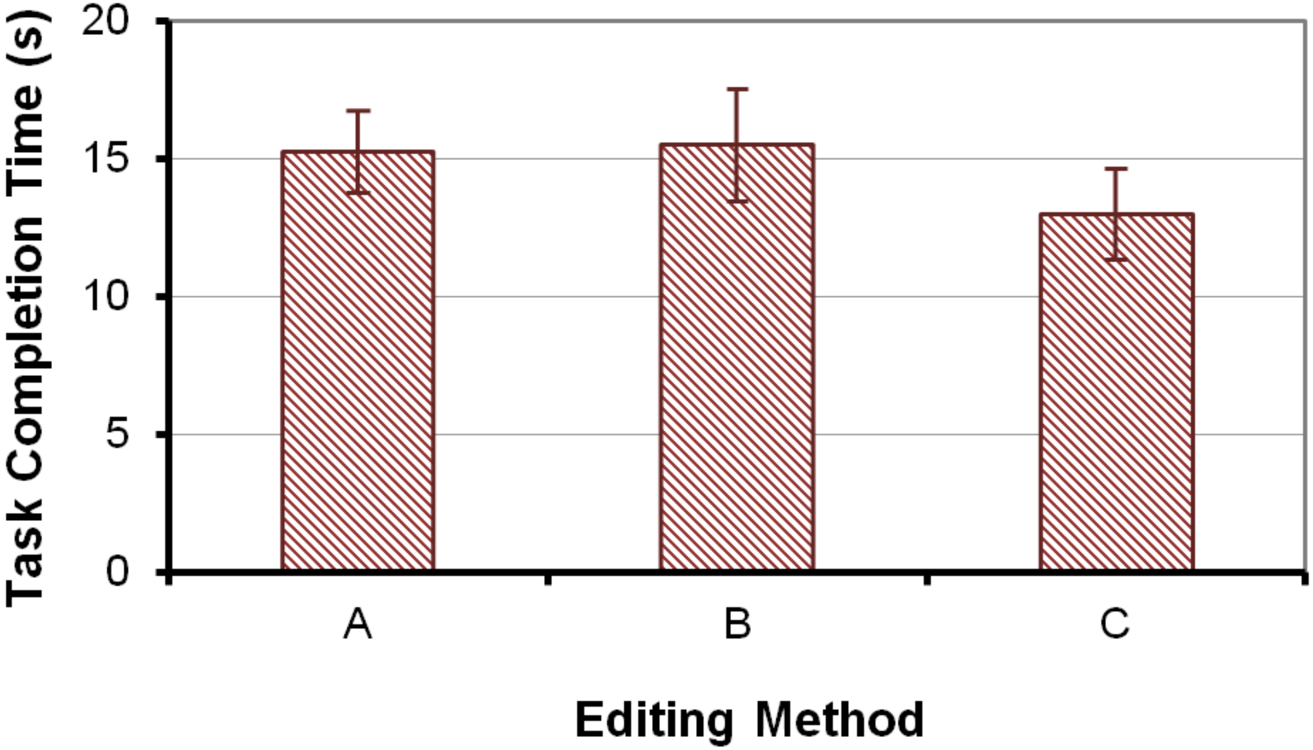
order: BCA

order: CAB

Group effect is small.
Counterbalancing worked!

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Results – Chart



MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

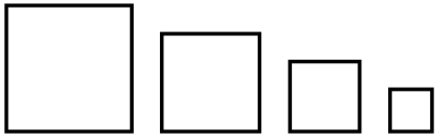
Other Techniques

- Instead of using a Latin square, all orders (n!) can be used
  - Preferable if possible
- Conditions can be randomized
  - Randomizing best if the tasks are brief and repeated often (examples below)

all orders of A, B, C:

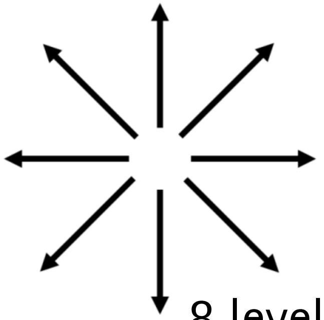
A	B	C
A	C	B
B	C	A
B	A	C
C	A	B
C	B	A

Target size



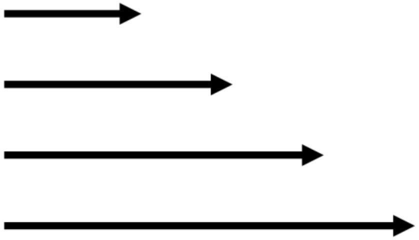
4 levels

Movement direction



8 levels

Movement distance



4 levels

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

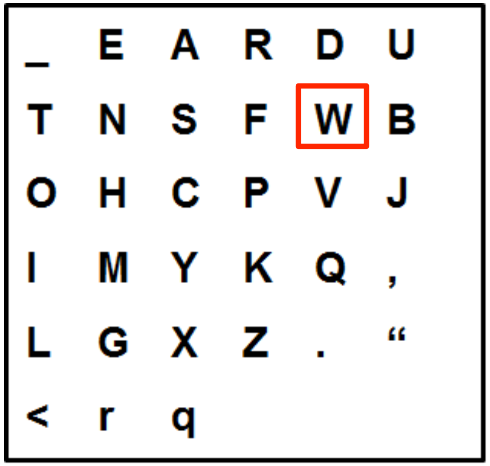
Asymmetric Skill Transfer

- Asymmetric skill transfer: Differences in the amount of improvement depending on the order of testing
  - Sometimes occurs in within-subjects designs
- There may be different learning effects for A → B than for B → A
  - In this case there is a group effect
- Example (next slides): Scanning keyboard
  - Keyboard image + hardware button, keys are highlighted one-by-one (scanned), user presses hardware button when desired key highlighted
  - With and without word prediction
  - Within-subjects design

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

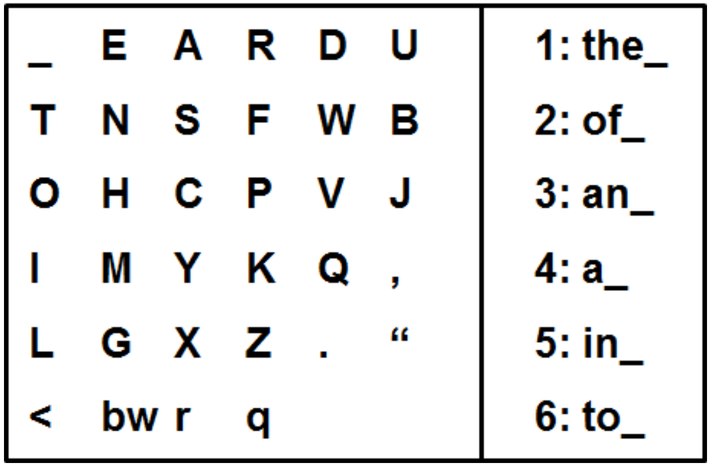
Results

letters only keyboard

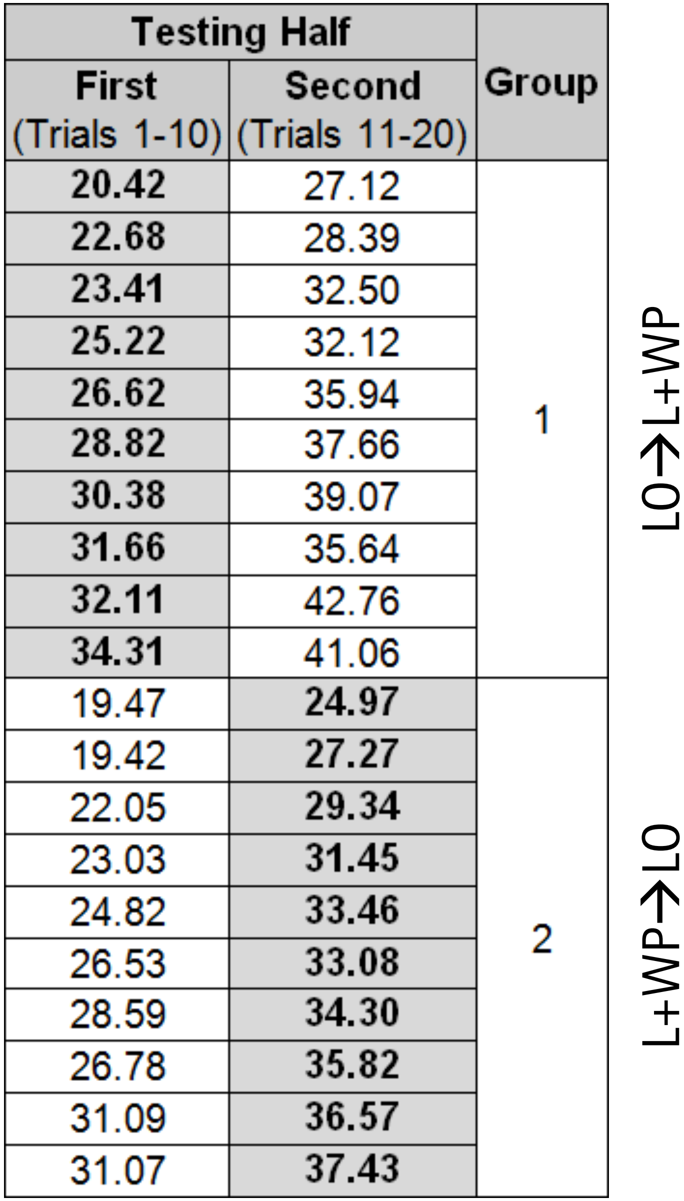


_	E	A	R	D	U
T	N	S	F	W	B
O	H	C	P	V	J
I	M	Y	K	Q	,
L	G	X	Z	.	“
<	r	q			

letters and word prediction keyboard



_	E	A	R	D	U	1: the_
T	N	S	F	W	B	2: of_
O	H	C	P	V	J	3: an_
I	M	Y	K	Q	,	4: a_
L	G	X	Z	.	“	5: in_
<	bw r	q				6: to_

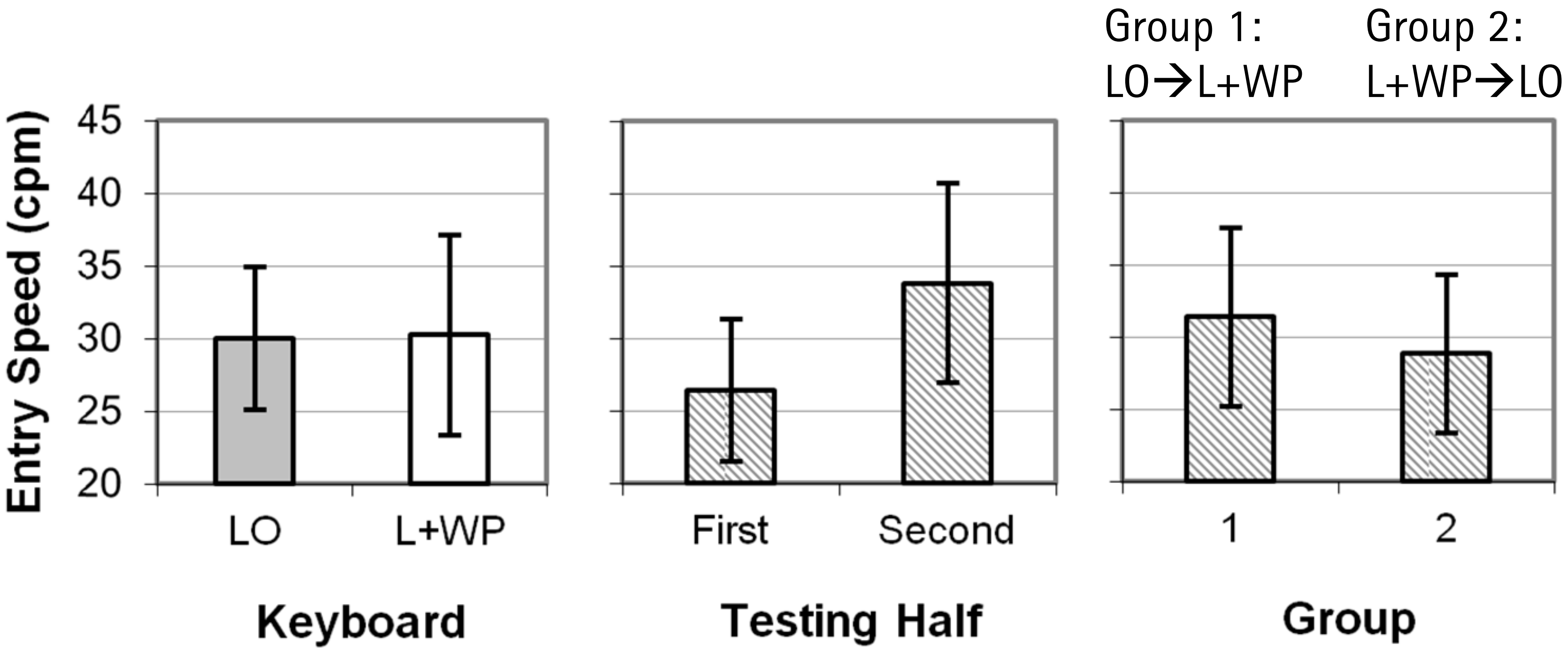


Testing Half		Group	
First (Trials 1-10)	Second (Trials 11-20)		
20.42	27.12	1	LO→L+WP
22.68	28.39		
23.41	32.50		
25.22	32.12		
26.62	35.94		
28.82	37.66		
30.38	39.07		
31.66	35.64		
32.11	42.76		
34.31	41.06		
19.47	24.97	2	L+WP→LO
19.42	27.27		
22.05	29.34		
23.03	31.45		
24.82	33.46		
26.53	33.08		
28.59	34.30		
26.78	35.82		
31.09	36.57		
31.07	37.43		

▨ = letters only

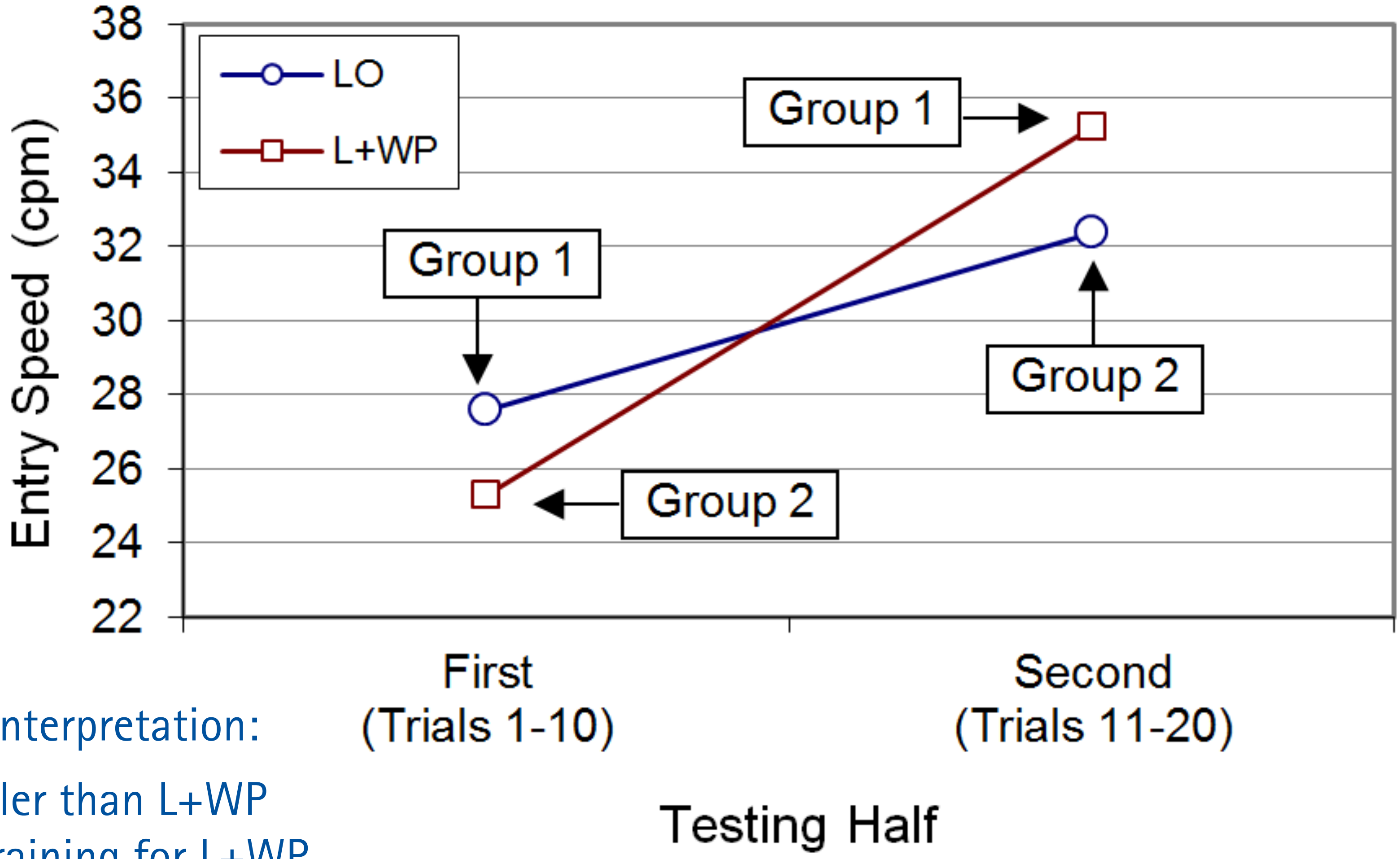
MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Results Summarized



MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Results, Asymmetric Skill Transfer



Plausible interpretation:

LO is simpler than L+WP

LO good training for L+WP

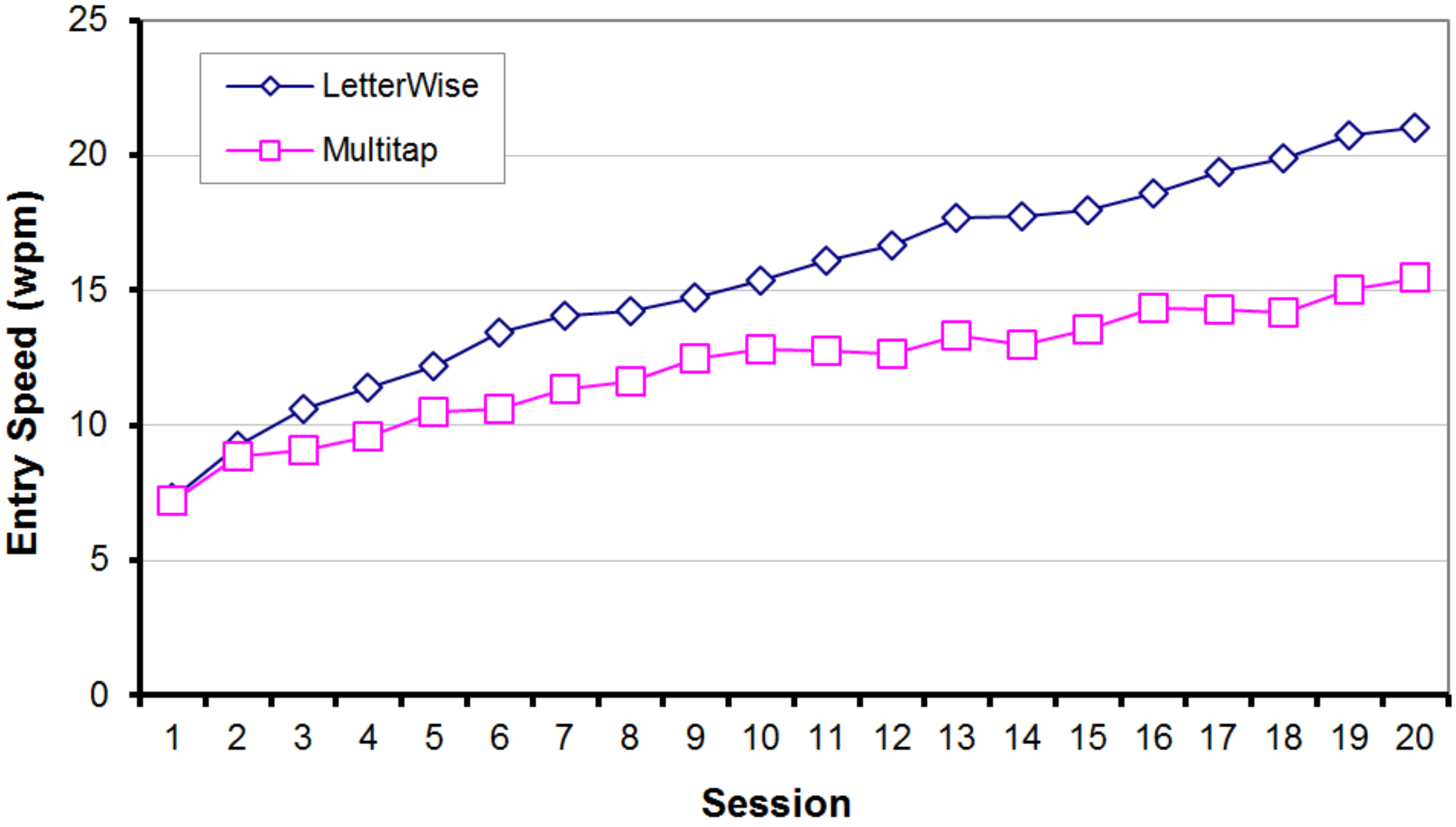
MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Longitudinal Studies

- Sometimes instead of "balancing out" learning effects, we want to to study learning effects
- If so, a longitudinal study is conducted
- IV: Amount of practice
- Participants practice over a prolonged period of time
- Practice units: blocks, sessions, hours, days, etc.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Longitudinal Study – Results[1]

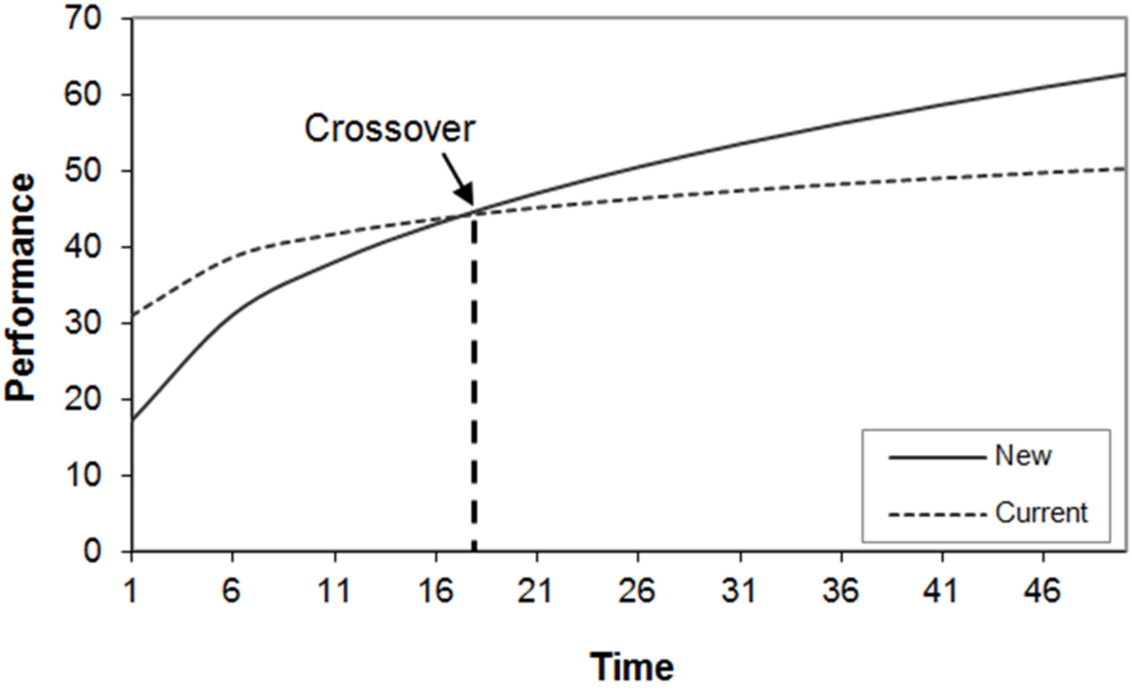


[1] MacKenzie, Kober, Smith, Jones, Skepner. LetterWise: Prefix-based disambiguation for mobile text entry. Proc. UIST 2001.

MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

The New vs. The Old

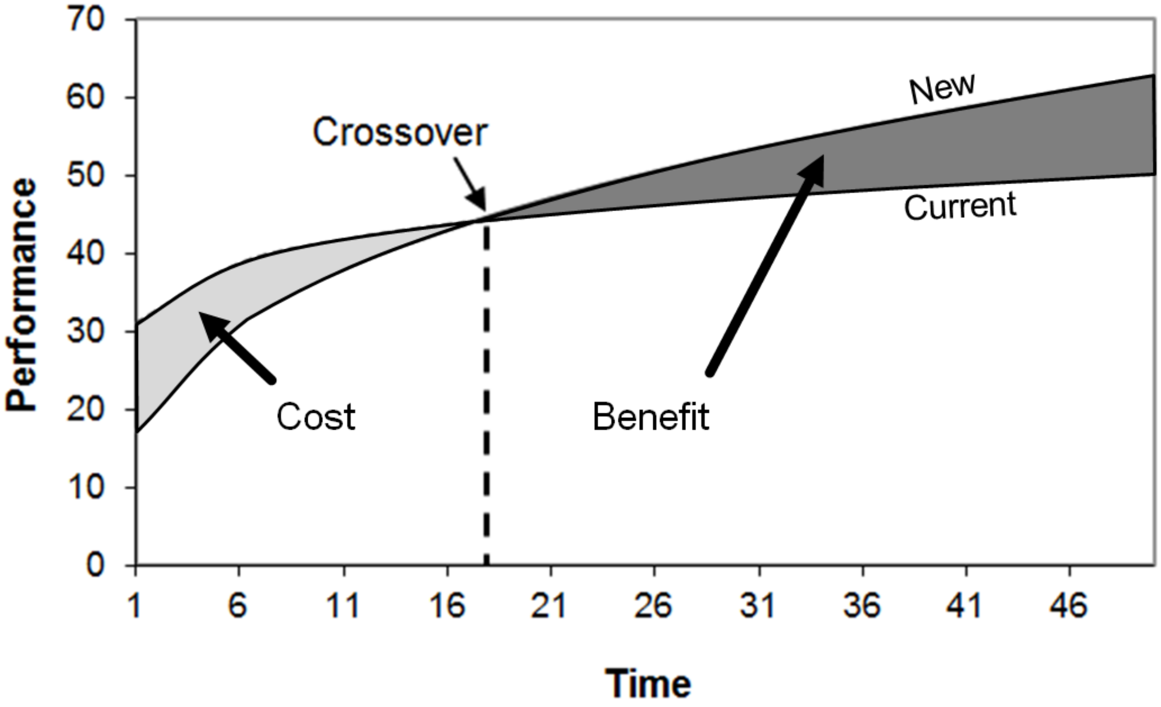
- Sometimes a new technique will initially perform poorly in comparison to an established technique
- A longitudinal study will determine if a crossover point occurs and, if so, after how much practice



MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.

Cost-Benefit Trade-Offs

- New, improved techniques sometimes languish
- Evidently, the benefit in the new technique is insufficient to overcome the cost in learning it



MacKenzie: Human-Computer Interaction - An Empirical Research Perspective.