Quantitative Research

Ahmed Sabbir Arif
University of California, Merced
https://www.theiilab.com/

Lecture notes in this series are based on
- Ahmed Sabbir Arif. 2021. Statistical Grounding. *Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice*, ACM
- Ann Blandford, Dominic Furniss, Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. Morgan & Claypool
- Jonathan Lazar, Jinjuan Feng, Harry Hochheiser. 2017. Research Methods in Human-Computer Interaction. Morgan Kaufmann
- I. Scott MacKenzie. 2013. Human-Computer Interaction: An Empirical Research Perspective, Morgan Kaufmann
- Interaction Design Foundation. 2022. Design Thinking
- Lecture notes of Amy Bruckman, Mark Dunlop, Niels Henze, I. Scott MacKenzie, Laura Moody, Albrecht Schmidt, Kami Vaniea

inclusive interaction lab

1

Quantitative Research

- Quantitative research involves the process of objectively collecting and analyzing numerical data to describe, predict, or control variables of interest
  - The goals are to test causal relationships between variables, make predictions, and generalize results to wider populations
- Methods: experimental, correlational

Saul McLeod. 2019. Qualitative vs Quantitative Research. Simply Psychology.

2

inclusive interaction lab

2

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

inclusive interaction lab

3

3

Internal Validity

- Definition: The extent to which the effects observed are due to the *test conditions*, such as Qwerty vs. new keyboard
- 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

inclusive interaction lab

4

4

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

inclusive interaction lab

5

5

Internal Validity: 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?

inclusive interaction lab

6

6

Internal Validity: 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 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?

inclusive interaction lab

7

7

Internal–External Validity 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, fiddling, or secondary tasks

inclusive interaction lab

8

8

Relationships: Circumstantial & Causal

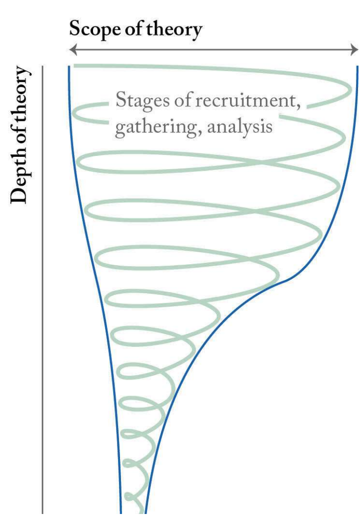
- As noted earlier:
  - Correlational methods → circumstantial relationships
  - Experimental methods → causal relationships
- Causal-and-effect conclusions not possible if the independent variable is a *naturally occurring attribute* of participants (e.g., gender, personality type, handedness, first language, political viewpoint)
- These attributes are legitimate independent variables
- But, they cannot be assigned to participants; thus causal relationships not valid

9

inclusive interaction lab

9

Grounded Theory (Recap)



- Grounded theory involves constructing hypotheses and theories through the process of collecting and analysis data
- The scope of the theory generated starts off broad, narrowing as we learn more

10

inclusive interaction lab

10

Grounded Theory: Example

- Find a storyline and characters for a future adventure game that will appeal to the target user group
  - Greek myth?
  - Gods, demigods, mortals?
  - Which demigods?
  - How should they look?
  - …
  - How about a game based on Norse myth?

the thor we thought we gonna get

the thor we got

Matthew Byrd. 2021. God of War Ragnarok: "Fat Thor" Design Divides the Internet Despite Being Perfect. Den of Geek

inclusive interaction lab

11

11

Experiment Design

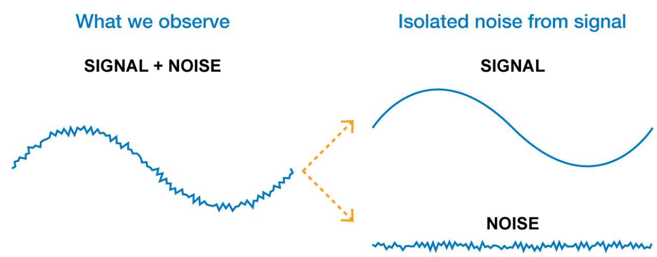
inclusive interaction lab

12

12

<u>Experiment Design:</u> Goal

- Signal and noise metaphor for experiment design:
  - Enhance the signal (a variable of interest), minimizing the noise (everything else)



13

inclusive interaction lab

13

Methodology

- *Methodology* is the way an experiment is designed & carried out

- Methodology is critical:
  - What methodology?
  - Don't just make it up because it seems reasonable

- What are the experimental variables?
  - Properly formed research questions inherently identify experimental variables

14

inclusive interaction lab

14

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 (i.e., 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

inclusive interaction lab

15

15

Independent Variable: 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)

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

inclusive interaction lab

16

16

Human Characteristics: Revisited

- Human characteristics are *naturally occurring attributes*
- Examples:
  - Gender, age, height, weight, handedness, grip strength, finger width, visual acuity, personality trait, political viewpoint, first language, shoe size, 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

17

inclusive interaction lab

17

Independent Variable: How Many?

- An experiment must have at least one independent variable
- Possible to have 2, 3, or more IVs
- 1 to 3 IVs are recommended at the most since the number of "effects" increases rapidly with the size of the experiment

Independent Variables	Effects					Total
	Main	2-way	3-way	4-way	5-way	
1	1	-	-	-	-	1
2	2	1	-	-	-	3
3	3	3	1	-	-	7
4	4	6	3	1	-	14
5	5	10	6	3	1	25

18

inclusive interaction lab

18

Dependent Variable

- A *dependent variable* is a measured human behaviour (related to an aspect of the interaction involving an independent variable)
- "Dependent" because it depends on what the participant does
- 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

19

inclusive interaction lab

19

Dependent Variable: Unique/Custom

- Any observable, measurable behaviour is a legitimate dependent variable (provided it has the potential to reveal differences among the test conditions)
- So, you can define your own dependent variables
- Example: *walking & texting*
  - Number of wrong turns
  - Number of collision
  - Incidents are manually logged
  - Clearly defined, thus reproducible

Ahmed Sabbir Arif, Benedikt Iltisberger, Wolfgang Stuerzlinger. 2011. Extending Mobile User Ambient Awareness for Nomadic Text Entry. The 23rd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OzCHI 2011). ACM, NY, 21-30.

20

inclusive interaction lab

20

Control Variable

- A *control variable* is a 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 keyboard background on text entry performance?
  - Independent variables: background
  - Dependent variable: entry speed (wpm), accuracy (er)
  - Control variables
    - Keyboard and key size,
    - Font family and size,
    - Position and posture, etc.

Ashish Yadav, Ahmed Sabbir Arif. 2018. Effects of Keyboard Background on Mobile Text Entry. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018). ACM, New York, NY, USA, 109-114.

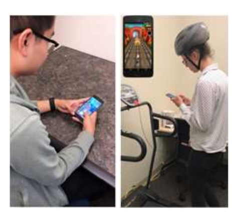
inclusive interaction lab

21

21

Random Variable

- A *random variable* is a 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 the game affects the performance an endless runner game differently in stationary and mobile settings?
  - Independent variable: game pace (slow, default, fast)
  - Dependent variable: attempts, score
  - Random variables
    - Prior experience playing endless runner games,
    - Prior experience playing *Subway Surfer*,
    - Amount of coffee consumed prior to testing, etc.



Mudit Misra, Elena Márquez Segura, Ahmed Sabbir Arif. 2019. Exploring the Pace of an Endless Runner Game in Stationary and Mobile Settings. The 2019 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY 2019 Extended Abstracts). ACM, NY, 543-550.

inclusive interaction lab

22

22

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 in the measured behaviours.
Control	Improves internal validity since variability due to a controlled circumstance is eliminated	Compromises external validity by limiting responses to specific situations and people.

inclusive interaction lab

23

23

Confounding Variable

- A *confounding variable* is a 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): small camera mounted on eyeglasses
    - Far camera (B): high-tech camera mounted above system display
  - Dependent variable: task completion time
  - But "camera" is a confounding variable: camera A for the near setup, camera B for the far setup
  - Are the effects due to camera distance or to some aspect of the different setups?

inclusive interaction lab

24

24

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 *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)

inclusive interaction lab

25

25

Experiment Task: Examples

- Usually the task is self-evident (follows the research idea)
- Research idea: a speech-based method to write source code
  - Experiment task: write a function using
    a) The speech-based method
    b) The conventional method (editor)
- Research idea: an haptic feedback method for joysticks
  - Experiment task: select targets
    a) With haptic feedback
    b) Without haptic feedback

inclusive interaction lab

26

26

<u>Experiment Task:</u> Knowledge-based Tasks

- Most experiment tasks are *performance-based* or *skill-based*
  - Creating an equation, programming a destination location, etc.
- Sometimes the task is *knowledge-based*
  - Using an Internet search interface to find *"the birth date of Albert Einstein"*
  - In this case, participants become contaminated (in a sense) after the first run of task, since they have acquired the knowledge
  - Experimentally, this poses <u>**problems**</u>
- A creative approach is needed
  - For the other test condition, slightly change the task: *"…of William Shakespeare"*

inclusive interaction lab

27

27

Experiment Procedure

- The *procedure* encompasses everything occurs with participants
- Includes the experiment task, and everything else…
  - Arriving, welcoming
  - Signing a consent form
  - Instructions given to participants about the experiment task
  - Demonstration trials, practice trials
  - Rest, breaks
  - Administering of a questionnaire or an interview

inclusive interaction lab

28

28

Instructions

- Very important (best to prepare in advance; write out)
- Often the goal in the experiment task is *"to proceed as quickly and accurately as possible but at a pace that is comfortable"*
- Other instructions are fine, as per the goal of the experiment or the nature of the tasks, but…
  - Give the 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

inclusive interaction lab

29

29

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

inclusive interaction lab

30

30

Participants: Sampling Procedures

- Purposive sampling (aka judgement sampling)
  - Selecting a sample (of users) who most likely to address the research question efficiently
- Theoretical sampling (advocated within grounded theory)
  - Recruiting people who are most likely to test, expand, and help build the theory that is emerging through data gathering and analysis
  - Only possible when data collection and analysis are interleaved
- Convenience sampling
  - Working with the most accessible participants (the easiest approach)
- Snowball sampling
  - Each participant introduces further participants to the study

inclusive interaction lab

31

31

Participants: How Many?

- With too few experimental effects fail to achieve statistical significance
- With too many statistical significance for effects of no practical value
- Use the same number of participants as used in similar research

Ahmed Sabbir Arif. 2021. Statistical Grounding. *Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice (1st ed.)*. ACM, New York, NY, USA, 59–99.

inclusive interaction lab

32

32

Questionnaires

- Questionnaires are used in most HCI experiments
- Two purposes
  - 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
- Questions constructed according to how the information will be used

inclusive interaction lab

33

33

Questionnaires: Examples

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

inclusive interaction lab

34

34

Questionnaires: Participant Feedback

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

NASA-TLX Like

Eye fatigue:

1	2	3	4	5	6	7
Very high						Very low

ISO 9241-9

inclusive interaction lab

35

35

Experiment Design: Within vs. Between-subjects

- Two ways to assign conditions to participants:
  - *Within-subjects:* each participant is tested on each condition (aka *repeated measures*)
  - *Between-subjects:* each participant is tested on one condition only

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

inclusive interaction lab

36

36

Experiment Design: Within vs. Between-subjects

- 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)

iil inclusive interaction lab

37

37

Experiment Design: Counterbalancing

- Order effects are 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 a Latin square
  - Distinguishing property of a Latin square → each condition occurs precisely once in each row and column (next slide)

iil inclusive interaction lab

38

38

Counterbalancing: 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

inclusive interaction lab

39

39

Counterbalancing: 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, $n-2$, …

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

inclusive interaction lab

40

40

Counterbalancing: Example

- A researcher seeks to determine if three editing methods (A, B, C) differ in the amount of time to do a common editing task

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

- Conditions are assigned within-subjects
- 12 participants are divided into 3 groups (4 per group)
- Methods administered using a 3 × 3 Latin Square

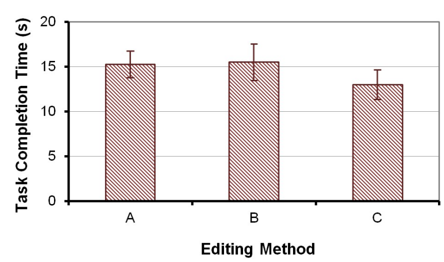
inclusive interaction lab

41

41

Counterbalancing: Example (Results)

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			



Group effect is small
∴ Counterbalancing worked!

inclusive interaction lab

42

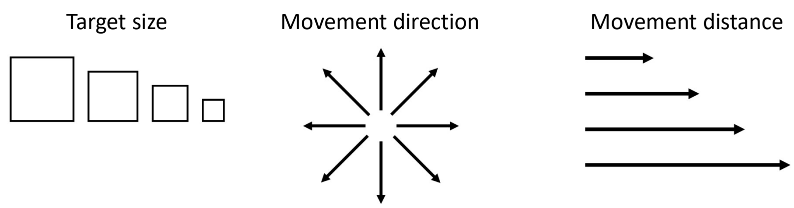
42

Counterbalancing: Other Techniques

- Instead Latin square, all orders (n!) can be used
- Conditions can be randomized
  - Randomizing best if the tasks are brief and repeated often

All orders
3 × 3

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



inclusive interaction lab

43

Asymmetric Skill Transfer

- Asymmetric skill transfer is an interference that occurs in within-subjects designs
  - It occurs when the learning of new information or behaviors interacts with previous learning or memories and interferes with the acquisition of the new information

inclusive interaction lab

44

Asymmetric Skill Transfer: Example

_	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			

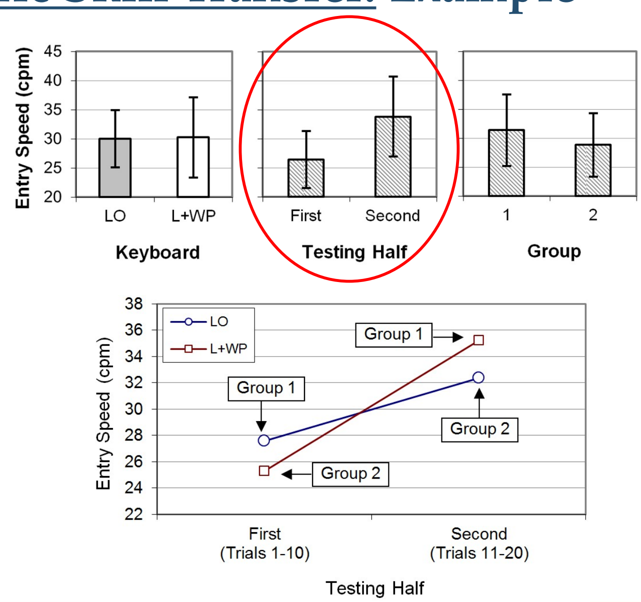
Letter (LO)

_	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_

Letter + Word Prediction (L+WP)

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

Asymmetric Skill Transfer: Example



Longitudinal Studies

- Sometimes instead of *balancing out* learning effects, the research seeks to promote and investigate learning
- If so, a *longitudinal study* is conducted
- *Practice* is the independent variable
- Participants are practiced over a prolonged period of time
- Practice units: blocks, sessions, hours, days, etc.

inclusive interaction lab

47

47

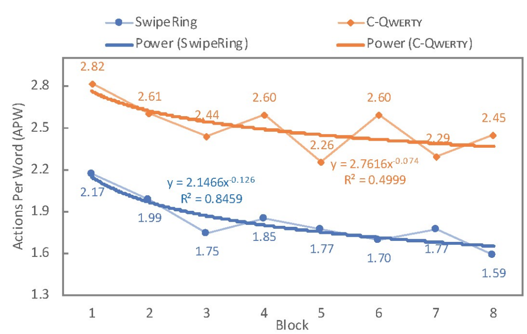
Recap: Power Law of Learning

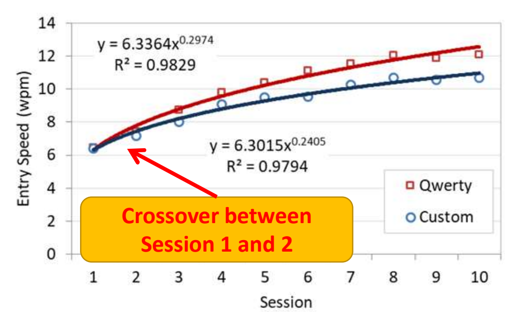
$$T_n = T_1 \times n^a$$

Predicting Time

$$S_n = S_1 \times n^a$$

Predicting Speed





Gulnar Rakhmetulla, Ahmed Sabbir Arif, Steven J. Castellucci, I. Scott MacKenzie, Caitlyn Seim. 2021. Using Action-Level Metrics to Report the Performance of Multi-Step Keyboards. Graphics Interface Conference (GI 2021). Article 15, 127-137.

Steven J. Castellucci, I. Scott MacKenzie, Mudit Misra, Laxmi Pandey, Ahmed Sabbir Arif. 2019. TiltWriter: Design and Evaluation of a No-touch Tilt-based Text Entry Method for Handheld Devices. International Conference on Mobile and Ubiquitous Multimedia (MUM 2019). ACM, Article 7, 8 pages.

inclusive interaction lab

48

48