CMPT 363: User Interface Design
Summer 2021

Week 10: Analytical Evaluation: Cognitive Walkthrough & Fitts' Law

Instructor: Victor Cheung, PhD

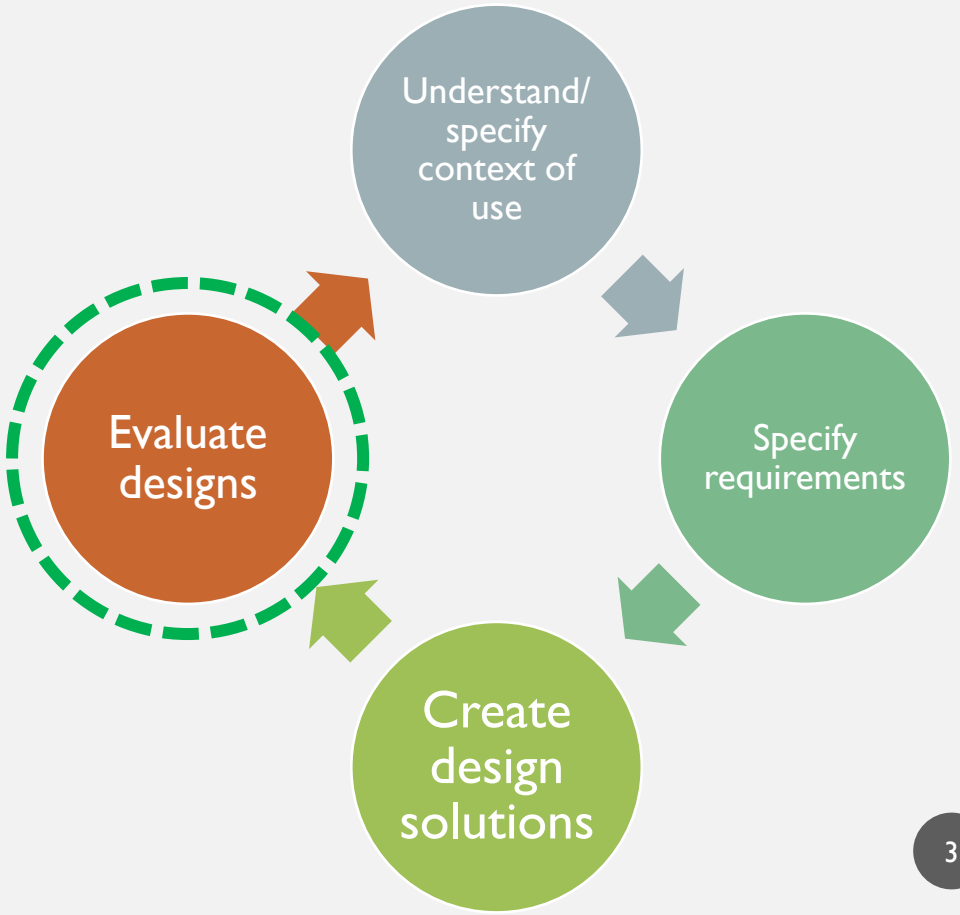
School of Computing Science, Simon Fraser University

Recap from Last Lecture

- Analytical evaluation
  - Cognitive walkthrough
    - Parts involved – prototype, scenario, action, target users
    - Focus – first time usage without prior training
  - Fitts' Law
    - What it is – began with measuring difficulty in reach a target, adapted to predicting movement time
    - Examples of use – commonly used in interface design to speed up selection

Today

- Analytical evaluation
  - GOMS & KLS
  - Involving users implicitly



GOMS & KLM

Other predictive models

GOMS

- A cognitive framework describing tasks by Card, Moran, & Newell (1983)*
  - Goals – main objective of a task or the state the user wants to achieve
    - Examples: buy the id-book at an online bookstore, bake a cake
  - Operators – atomic cognitive & physical actions required to attain the goal (or sub-goals)
    - Examples: enter the book name at the online book store, click the buy button, mix the ingredients, put the mix into the oven
  - Methods – learned procedures for accomplishing the goal
    - Example: type the book name in the search field with a keyboard, click the search button, and then click the buy button under it
  - Selection rules – specification of which method should be used to attain the goal
    - Example: press the enter key to search or click the search button to search

*Card, Stuart, Moran, Thomas P., and Newell, Allen, The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, NJ (1983).

GOMS Model of Deleting a Word in A Sentence Using A Word Processor (MS Word)

- Goal: delete a word (not) in a sentence (I do not like drinking coffee.)
- Operators: recall/find word, move cursor, left mouse click, left mouse hold/release, locate/press key

Method 1: achieve goal by using the backspace key
- Recall/find where the word is in the sentence
- Position the cursor after the last character of that word
- Locate the backspace key on keyboard
- Press the backspace key until the whole word is deleted

Method 2: achieve goal by highlighting the word and deleting it
- Recall/find where the word is in the sentence
- Hold the left mouse button and drag the cursor overall the characters of that word, release when done
- Locate the backspace key on keyboard
- Press the backspace key once and delete the word

- Selection rules: if the word is short, use method 1, otherwise use method 2

KLM (Keystroke-Level Model)

- A simplified version of GOMS targeting mouse & keyboard operations by Card & Moran (1980)*
- Provides numerical predictions of user performance (time required) with different methods for the same task
  - Proposed a number of core times in performing physical actions with mouse & keyboard based on empirical studies

Operator name	Description	Time (s)
K	Pressing a single key or button Skilled typist (55 wpm) Average typist (40 wpm) User unfamiliar with the keyboard Pressing shift or control key	0.35 (average) 0.22 0.28 1.20 0.08
P	Pointing with a mouse or other device to a target on a display	1.10
P_1	Clicking the mouse or similar device	0.20
H	Homing hands on the keyboard or other device	0.40
D	Draw a line using a mouse	Variable depending on the length of line
M	Mentally prepare to do something, e.g. make a decision	1.35
R(t)	System response time – counted only if it causes the user to wait when carrying out his/her task	*t*

These values could change for different hardware setup and user groups – need to verify

From Ch. 15.4.2 of ID-Book (2011 edition)

*Card, Stuart, Morn, Thomas P., and Newell, Allen, The keystroke-level model for user performance with interactive systems, *Communications of the ACM*, 23 (1980), 396-210

KLM of Inserting a Word in A Sentence Using A Word Processor (MS Word)

- Predicted time of execution is a combination of mental & physical steps
- Example: insert the word "not" into the sentence "Running through the streets naked is normal."

Mentally prepare (M)	1.35
Reach for the mouse (H)	0.40
Position mouse before the word 'normal' (P)	1.10
Click mouse (P_1)	0.20
Move hands to home position on keys (H)	0.40
Mentally prepare (M)	1.35
Type 'n' (good typist) (K)	0.22
Type 'o' (K)	0.22
Type 't' (K)	0.22
Type 'space' (K)	0.22
Total predicted time:	5.68 seconds

One might argue that M could also happen between typing keys (need to make sure spelling is correct) – some testing, adjustments, and assumptions have to be made

From Ch. 15.4.2 of ID-Book (2011 edition)

Strengths & Weaknesses of Predictive Models

- Strengths
  - Provide a way of evaluating products or designs without directly involving users, thus takes less time and costs less
  - Enable evaluation when it is difficult to recruit participants to a usability study (e.g., motor impairments)
  - Different models available to different interface types
- Weaknesses
  - Measurements limited to products or designs with predictable tasks (e.g., doing a fixed set of actions, not creative activities)
  - Does not provide insights in qualitative aspects, such as learnability, aesthetics, and satisfaction
  - Based on expert error-free behaviour (i.e., does not take user error and confusion into consideration)

Involving Users Implicitly

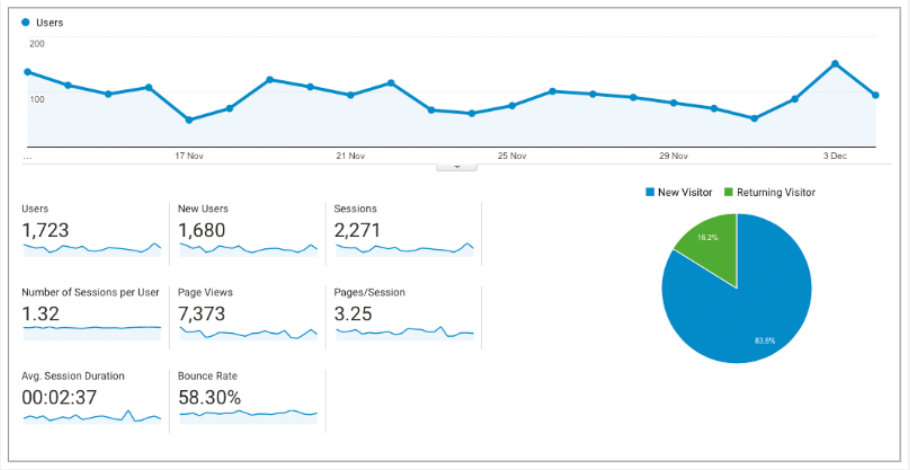
Web Analytics & A/B Testing

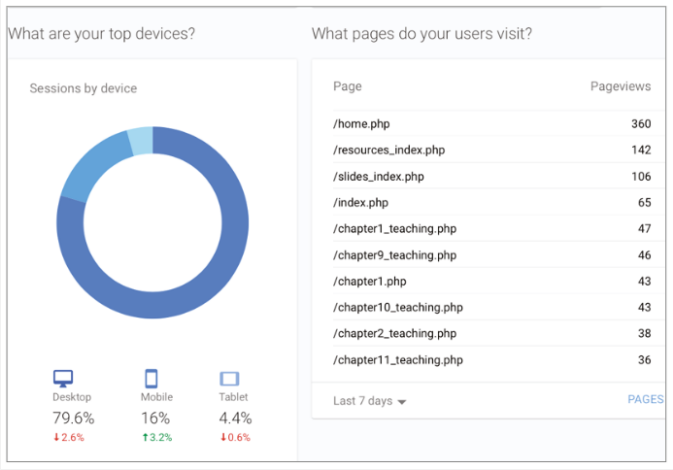
Objectives

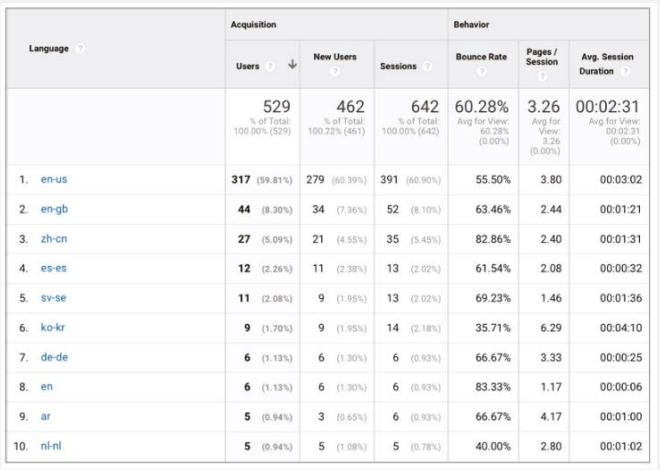
- To gather information about user activities/behaviour in using an interface in an unobtrusive way
  - Users can use the interface normally without any interruption
  - Can be done automatically so collection can be easy and in large scale
- Typical methods of collection & data being collected
  - Key logging (keys being typed)
  - Mouse tracking (movement of mouse, dwell time of mouse)
  - Heat map (where most clicks are, where most people spent looking)
  - Time spent (for how long does each page keep the viewer, how long does each step takes the user)

Web Analytics

- Specifically designed to analyze users' activity and behaviour on websites
  - Typically achieved by injecting some scripts on each webpage
- Useful to see a "big-picture" overview of user interaction (interface usage), interpretation is left for experimenter







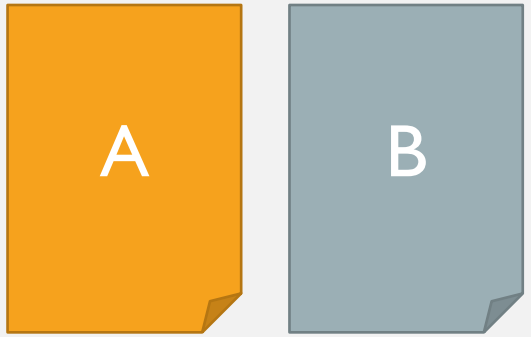
Language	Acquisition			Behavior		
	Users	New Users	Sessions	Bounce Rate	Pages / Session	Avg. Session Duration
	529 % of Total: 100.00% (529)	462 % of Total: 100.22% (461)	642 % of Total: 100.00% (642)	60.28% Avg for View: 60.28% (0.00%)	3.26 Avg for View: 3.26 (0.00%)	00:02:31 Avg for View: 00:02:31 (0.00%)
1. en-us	317 (59.81%)	279 (60.39%)	391 (60.90%)	55.50%	3.80	00:03:02
2. en-gb	44 (8.30%)	34 (7.36%)	52 (8.10%)	63.46%	2.44	00:01:21
3. zh-cn	27 (5.09%)	21 (4.55%)	35 (5.45%)	82.86%	2.40	00:01:31
4. es-es	12 (2.26%)	11 (2.38%)	13 (2.02%)	61.54%	2.08	00:00:32
5. sv-se	11 (2.08%)	9 (1.95%)	13 (2.02%)	69.23%	1.46	00:01:36
6. ko-kr	9 (1.70%)	9 (1.95%)	14 (2.18%)	35.71%	6.29	00:04:10
7. de-de	6 (1.13%)	6 (1.30%)	6 (0.93%)	66.67%	3.33	00:00:25
8. en	6 (1.13%)	6 (1.30%)	6 (0.93%)	83.33%	1.17	00:00:06
9. ar	5 (0.94%)	3 (0.65%)	6 (0.93%)	66.67%	4.17	00:01:00
10. nl-nl	5 (0.94%)	5 (1.08%)	5 (0.78%)	40.00%	2.80	00:01:02

Some analytics about the id-book.com website generated by Google Analytics in December 2018. Source: ID-Book Figure 16.5

A/B Testing

- Have two different designs (A & B), serving one group of users with design A, and the other group with design B

- Typically used to compare changes in a design
  - Examples: positions of buttons, use of terminologies
- Typically done in a large scale to produce meaningful results
- Combined with other evaluation methods (e.g., web analytics)



Steps to Conduct An A/B Test

1. Determine what to test (based on observation and design) and what to measure (based on goals)
2. Build the designs (usually 2, but can be more)
   - Purest form is to change only one "variable" (e.g., size of a button, use of a different word), but takes the most time
3. Determine number of participants, typically in the order of a few dozens
   - There are formulas to calculate that to guarantee a level of confidence, but in general the more the better
4. Eliminate as many confounding factors (outside circumstances that impact your test) as possible
   - Run the test at different times, over a longer period, with diverse group of participants, …etc.
5. Collect measurements & analyze, possibly repeat

Strengths & Weaknesses of Web Analytics & A/B Testing

- Strengths
  - Relatively easy to deploy, bigger sample size
  - Doesn't require giving users specific tasks to carry out
- Weaknesses
  - Don't really know what's happening with the users (e.g., cannot really telling what causes an error)
    - Results are only valid for the elements being tested (cannot make any prediction)
  - Requires a working interface to be deployed
  - Privacy issues (tracking might be done without people's knowledge)

A Modern Way to Get Users

- Crowdsourcing (Mechanical Turk)
  - Experimenter posts tasks as jobs (Human Intelligence Tasks, HITs), users complete tasks as turkers in return for rewards
- Strengths
  - affordable
  - reach more users with diverse demographics, 24x7
- Weaknesses
  - typically only works well with simple & definable tasks
  - quality of results might not be guaranteed
  - unintentional biased demographics due to access
  - designs are made public

https://www.mturk.com/

Summary

- Analytical evaluation
  - Cognitive walkthrough
    - Steps involved, its focus
  - Fitts' Law
    - What it is, examples of use
  - GOMS & KLM
    - What they are, strength & weaknesses
  - Involving users implicitly
    - What they are, strength & weaknesses

Post-Lecture Activity

- Read/watch these (and those in the slides)
  - ID-Book Ch. 16.2.2
  - Literature on Cognitive Walkthrough (8min read)
    https://www.interaction-design.org/literature/topics/cognitive-walkthrough
  - What is Fitts's Law (and some of its applications) by Denis Nguyen (5min watch)
    https://vimeo.com/200436363
  - Putting A/B Testing in Its Place
    https://www.nngroup.com/articles/putting-ab-testing-in-its-place/

- Exercise
  - Watch the following cognitive walkthrough example and identify what task it was testing, and what steps are involved
    https://www.youtube.com/watch?v=Op9huZ_GNAk
  - Use the KLM to model the 2 methods in p8 in deleting a word