

CS 279 Assignment 2: Conduct an Experiment

Part 1: Implementation

Miriam Cha and Melih Elibol

September 16, 2014

Webpage URL

<http://50.56.219.32>

Please note that the webpage is best supported when a browser window is maximized. We have tested the site's compatibility with Chrome, Safari, and Firefox.

Log Description

We design our logs to keep track of interface (CommandMap shown in Figure 1 or Ribbon), number of tab switches (same: 0 or different: 1), selection time, number of incorrect clicks, and total number of clicks per a trial. As shown in Table 1, each of our participants is assigned to complete 90 trials per an interface. We will find support for H_1 by evaluating mean selection time for each interface. Then, we will find support for H_2 by computing mean selection time and mean error rate with respect to tab switch for each interface.

ID	Interface	Trial	# of tab switches	selection time	# of incorrect clicks	# of Total clicks
3234	Ribbon	1	0	1.42	0	1
3234	Ribbon
3234	Ribbon	90	1	2.4	2	3
3234	CM	1	0	1.2	0	1
3234	CM
3234	CM	90	0	1.1	0	1

Table 1: Example of log entries per a trial

Upon completion of each interface, participants complete NASA-TLX worksheet to assess workloads. Table 2 shows notional log entries for NASA-TLX. At the end of an experiment, participant chooses an interface for preference. Entry logs similar to Table 2 and Table 3 are gathered to find support for H_3 .

ID	Interface	Mental Demand	Physical Demand	Temporal Demand	Hard Work	Frustration
3234	Ribbon	3	2	2	4	2
3234	CM	2	1	2	2	1

Table 2: Example of log entries for NASA-TLX

ID	preference
3234	CM

Table 3: Example of a preference log entry per a participant

Design Characterization

- **Independent variables:**

Interface

- **Dependent variables:**

Selection time, error rate, NASA-TLX measures (mental demand, physical demand, temporal demand, hard work, frustration), and interface preference

- **Control variables:**

User’s familiarity in MS Word 2007, resolution, computer specs, and screen size

- **Random variables:**

Weather, temperature in the room, vision power of a participant, reaction rate of a participant, noise level, and usual versions of MS Word that a participant uses

- **Design modification:**

- *Design modifications from what’s described in the paper:* Number and gender partitions of participants, computer specifications, screen size, and ‘correct’ sound component in addition to ‘incorrect’ sound.

- *Design choices that weren’t clear in the paper that we had to figure out on your own:* Methods of pointing (mouse, trackball, touchpad, joystick?), mouse acceleration, display of time, presentation methods of an instruction (“as quickly and accurately as possible”), audible deep sound, definition of error rate (CM inherently has a smaller number of total clicks if we include ‘tab clicks’ into ‘total number of clicks.’ Therefore, for Ribbon, we only count ‘command target clicks’ into ‘total number of clicks’.)

Internal and External Validities

- **Threat to internal validity; Can we mitigate them?; How?**

If we were to perform experiments for Study 2 at anytime throughout a week, external environmental factors could threat internal validity. In order to mitigate potential threat to internal validity, we plan to perform experiments at the same time period throughout the week. Additionally, we will keep the same location, lighting, computer, mouse, keyboard, task, and presentation method of an instruction.

- **Threat to external validity; Can we mitigate them?; How?**

Simply performing the experimental task of clicking command targets may not generalize to the world outside the lab that tends to include performing other additional tasks (e.g., email, doc edits, web browsing); For the purpose of replicating Study 2, this concern cannot be mitigated. It would be ideal if we can assess the similarity of the behaviors between the lab and the field but it is beyond the scope of Study 2. However, we plan to conduct an experiment in a realistic environmental setting (e.g., MD office). In order to mitigate potential external validity on gender imbalance (in the original experiment, the proportion of male to female was 8 to 1), we plan to recruit male and female participants who are knowledge to MS Word with nearly equal numbers.

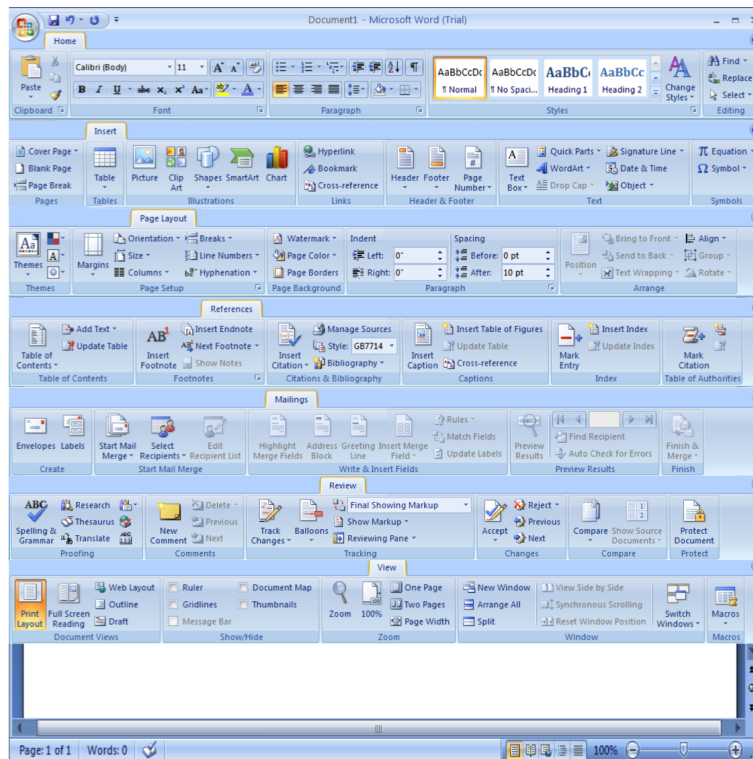


Figure 1: Our example CommandMap for Microsoft Word