Name/Student ID _____

Quiz 12 on Chapter 14 and the lectures

July 18, 2012

There are 4 questions, with a total value of 3.5 points; but the maximum score is 3 points. Please provide concise answers: Do not use more space than indicated. Good luck!

1. (1 point) In the lecture, we talked about the laboratory-based usability testing paradigm. 1) List the three most commonly used measures of *outcomes* in a task. 2) List three examples of *process* measures.

Slide 8 and following of July 10th; 14.2.1 in the book.

1) Any three of the following are fine, yielding ½ point:
   - task completion time
   - success/fail in completing the task
   - satisfaction in completing the task
   - evaluation of the usability of the design

2) Any three of the following are fine, yielding ½ point:
   - Total number of errors users encountered
   - Type of errors
   - Rate of errors
   - Prevalence of an error (number of users making the error)
   - Areas attended to during search
   - Think aloud / verbal protocol
   - Behavioral measures, such as keypresses, mouse movements and clicks, invoked commands
   - Subjective workload measures

2. (1.5 points) Identify and explain three likely *nuisance variables* (threats to validity) in the following study:

"Two interfaces – UI_1 and UI_2, the first by a contractor and the latter by its competitor – were compared in a usability test. The study was carried out in a cafeteria at the campus. Students were recruited as subjects from neighboring tables and asked to carry out two tasks. Task A was carried out, then task B, both on a laptop computer. The experimenter recorded performance with an HD camera. The experiment took two days: Thirty users were recruited during the first day to use UI_1. Thirty more were tested the following day for UI_2."

There is an extensive list in the slides of July 11th (20 and following, adopted from Cook & Campbell, 1979), and a subset of these were discussed in the lecture. I here repeat those that are relevant to the present case:

- **Sampling**: Students who appear at the cafeteria and volunteer for the study most likely represents a very biased sample of the general population.
- **Random incidents**: Although the number of subjects is relatively high for a usability study, it is likely that doing the study in a cafeteria might add random distractions and noise to the data. It is not therefore sure if even 30 users are enough.
- **Learning and other order effects**: Task A and B were carried out always in the same order, making it possible that users learned from A and were able to do B better.
  - The same criticism can be applied to testing of UI1 and UI2.

However, *the bigger point here is to be able to identify what may have gone wrong and might hamper the validity of the conclusions*. Thus, innovative other explanations should be considered. For example, the following are likely and should be credited:

- Guessing of the experiment's purpose, e.g. that UI1 is the competitor and UI2 is the company's own product.
- Evaluator effect: Who was the evaluator, a novice or an expert?
- Being tested in a cafeteria setting when fellow students witness your performance might change your behavior.
- Moreover, the laptop may have included software that the users are not familiar with. This might reduce the external validity of the results.

3. (1/2 point) Define the following terms: *Independent variable*, *dependent variable*

14.3 in the book; Slide 16 from July 10th

4. (1/2 point) Define the following terms: *Validity, observational field studies*

Validity: Slides 20 and following of July 11th

Observational field studies: Slides 35-36 of July 11th

Validity refers to the confidence with which we can conclude that the dependent variable caused the observed change in the independent variable. In the lecture we discussed four subtypes:

1. Statistical inference

Can we infer that the observed change is reliable?

2. Internal

Was it the IV that caused the observed change in the DV?

3. Construct

Is the IV really the cause of the DV?

4. External

Do the results generalize?

It is sufficient to give just the general definition.

Observational field studies refer to studies where the system/design under study was NOT introduced *by the evaluator*, but such existed independent of him/her. This allows the observation of natural interaction that is independent of the evaluator. (By contrast, experimental field studies always intervene somehow, either by introducing a system, setting up some conditions, or by eliminating (nuisance) factors.)