Introduction to Preparing, Designing, and Executing User Studies in HCI

By Cha Lee

Plagiarized from all over the internet. To all the other authors I have stolen from… I'm sorry!

Why is this Necessary?

- User studies are by their nature murky and produce subjective results.
  - Process is not understood well by engineers.
  - Many published results are questionable.
- Attempting to de-mystify the process.
- Attempting to explain the process using laymen's words.

User Study Phases

- Preparation
  - Dotting all the I's and crossing all the T's.
  - Getting the human subjects.
- Design
  - The important points you need to understand.
  - The different factors which affect the validity of your results.
- Execution
  - Guidelines for executing a user study.

Preparation

- Two simple steps:
  - Get approved to conduct human subject experiments.
  - Get your study approved.

Preparation: Getting Approved to Conduct Human Subject Experiments

- IRB Approval
  - Institutional Review Board (IRB): oversees and approves research involving humans.
  - Every institution has them (including us)!
  - UCSB has the Human Subjects Committee (HSC)
    - Mandated by Department of Health and Human Services
    - Responsible to the Office for Human Research Protections (OHRP)
    - HSC has stipulated certain rules and regulations which must be adhered to: http://www.research.ucsb.edu/compliance/humansubjects/regulations-policies-and-guidance

Preparation: Human Subjects Training

- 2001: National Institute of Health (NIH) began requiring mandatory training in ethics of human research for its researchers.
- UCSB also made this a campus wide requirement at this time.
  - All UCSB researchers must take an online training module: http://hstraining.orda.ucsb.edu

Preparation: ORAHS and Project Approval

- The study must also be approved by the HSC.
- Study proposals submitted to Office of Research Application for the use of Human Subjects (ORAHS):
  - Must be approved by faculty.
  - https://orahs.research.ucsb.edu/Main/Search/Search.aspx

Preparation: Acquiring Human Subjects

- Guideline: participants should not feel obligated, pressured, or trapped into participating either mentally, socially or physically.

- Volunteers
  - Small funds
  - Fewer participants and more noise in results
  - They may feel they are doing you a favor

- Paid Participants
  - Bread and butter of user studies
  - A paper trail must be established per University rules and departmental rules.
  - Always consult the financial office.

Preparation: Human Subject Sources

- Email Lists
  - Departmental email lists
  - Usually requires moderator approvals
- Off-campus Sources
  - Public boards (virtual or physical)
  - Craig's List
- Human Subject Pools
  - Preferred source.
  - SONA system run by Psychology Department

Design: The Process of Designing a User Study

- The most critical step!
  - Begins at identifying the problem
  - Ends when trials begin
- Often given insufficient thought!
  - Arrogance and apathy from engineers
  - Every design decision can have unintended consequences
- Three steps:
  - Understanding the problem in its entirety.
  - Identifying the correct question to ask.
  - Creating the hypothesis.

Design: Understanding the Problem in its Entirety

- Essential.
- Requires reading the literature.
  - Provides information on understood effects.
  - Reduces chances of investigating meaningless effects.
  - Provides guidance for reducing confounding factors.
- Requires first-hand experience with similar interfaces, systems, or applications.
  - Gives you a "feel" for what users will experience.
  - Provides an understanding of how non-central factors (e.g. ergonomics) may affect your study.

Design: Identifying the Question

- Science is about asking the right question.
- User studies require very specific and precise questions.
- Good example:
  - "Which interface allows users to select objects more accurately and quickly using a hand-held display?"
- Bad example:
  - "Which interface is better?"
  - Open ended.
  - Does not specify what "better" means.

Design: Creating the Hypothesis

- "Hypotheses are an expression in declarative form and bridges the gap between theory and empirical inquiry."
- REQUIRED for user experiments BEFORE trials begin.
- Provides a guideline and direction to the inquiry.
- *Without a proper set of hypothesis, the study is nothing but exploratory work and should not be reported as valid findings.*

Design: Experimental Validity

- **Takes experience to truly understand.**
- 5 different factors to validity:

1. Statistical conclusion validity:
   - The degree to which conclusions reached about causal relationships are justified.
   - What is the statistical power of your results?
   - How much can you really say about your results?

2. Intentional validity:
   - The extent to which the chosen constructs (task or test) and measures adequately assess what the study intended to study.
   - Did you fairly design the experiment?

3. Ecological validity:
   - The extent to which research results can be applied to real life situations outside of research settings.
   - How realistic does your experiment replicate real world conditions?

Design: Experimental Validity

4. Internal validity:
   - "An inductive estimate of the degree to which conclusions about causal relationships can be made (e.g. cause and effect), based on the measures used, the research setting, and the whole research design."
   - Confounding variables:
     - **History**: the specific events occurring between the first and second measurements in addition to the experimental variables.
     - **Maturation**: processes within the participants as a function of the passage of time (not specific to particular events), e.g., growing older, hungrier, more tired, and so on.
     - **Testing**: the effects of taking a test upon the scores of a second testing.
     - **Instrumentation**: changes in calibration of a measurement tool or changes in the observers or scorers may produce changes in the obtained measurements.
     - **Statistical regression**: operating where groups have been selected on the basis of their extreme scores.
     - **Selection**: biases resulting from differential selection of respondents for the comparison groups.
     - **Experimental mortality**: or differential loss of respondents from the comparison groups.
     - **Selection-maturation interaction**: e.g., in multiple-group quasi-experimental designs

Design: Experimental Validity

5. External validity:
   - What external factors may indicate your results are not generalizable?
   - Confounding variables:
     - **Aptitude-Treatment-Interaction**: The sample may have certain features that may interact with the independent variable, limiting generalizability.
     - **Situation**: All situational specifics (e.g. treatment conditions, time, location, lighting, noise, treatment administration, investigator, timing, scope and extent of measurement, etc. etc.) of a study potentially limit generalizability.
     - **Pre-Test Effects**: If cause-effect relationships can only be found when pre-tests are carried out, then this also limits the generality of the findings.
     - **Post-Test Effects**: If cause-effect relationships can only be found when post-tests are carried out, then this also limits the generality of the findings.
     - **Reactivity (Placebo, Novelty, and Hawthorne Effects):** If cause effect relationships are found they might not be generalizable to other settings or situations if the effects found only occurred as an effect of studying the situation.
     - **Rosenthal Effects:** Inferences about cause-consequence relationships may not be generalizable to other investigators or researchers.

Execution

- The easiest, but most fatiguing part!
- Many steps, depending on specific situations:
  - Automation
  - Standard operating procedures
  - Team work
  - The power of observation
  - Data analysis
  - Interpreting results

Execution: Automation

- Minimize your work!
- When gathering data (demographic or measurements), reduce manual entry!
- Record EVERYTHING
- Use coded triggers for multi-step tasks:
  - Provides higher accuracy
  - Removes user error and researcher error

Execution: Standard Operating Procedures

- Be thorough and precise in your procedures!
- Have a set procedure for everything:
  - What you say to each subject.
  - How session breaks are handled.
  - How equipment is worn, taken off, etc.
  - Etc.
- Follow the procedures!
  - Minimizes the effect you (the researcher) may have on the results.
  - Minimizes the effect of malicious users.
    - Users will lie and cheat! *Even when they don't mean to.*

Execution: Team Work

- When possible, use multiple study administrators.
  - Reduces fatigue.
  - Encourages accuracy and consistency.

- Cable wranglers are crucial!

Execution: The Power of Observation

- Two types of scientific research: passive (observational) and active (experimental)

- Even with active research, careful observation is critical:
  - May discover new effects not discernable in the measured results.
  - May discover new behavioral patterns in users which were unpredicted.
  - May discover artifacts in the study.

Execution: Data Analysis

- Experimental design will dictate what type of statistical analysis to use:
  - T-Test
  - Confidence Interval Test
  - ANOVA:
    - Within or Between Subjects
    - One-way, two-way, multiple-way
  - Many more complex designs, way too much for this document to cover.

- Statistics Department provides help with their lab:
  - http://www.pstat.ucsb.edu/statlab.htm

Execution: Interpreting Results

- Understand the effect of your population on your results:
  - What were your users like?
  - Age, ethnic background, education, physical attributes, etc.

- No such thing as VERY SIGNIFICANT!!!
  - Statistical results only imply significance or non-significance.
  - All significance values are *arbitrary*, although the "standard" is 95%

- You did not find an effect **does not mean** there is no effect.

Conclusion

- Guidelines:
  - A user experiment uses empirical data to prove a theory; it does not try to create one.
  - Always focus on the question you are asking, and don't get sidetracked in your experiment.
  - Don't try to test many things at once.
  - Understand your target population.
  - Pilot studies are part of the design phase!
  - Choose a meaningful analysis.

- If you ask Cha for help, put him as an author… or NO SOUP FOR YOU!