Today's Agenda

- Evaluation methods
  - User testing
  - Experiments

Announcement: Quiz #5

Quiz # 5

- Thursday, Nov. 2 in class
- Via Blackboard – Bring your laptop to class!
- Open book and open notes

Individual Assignment-Case Study (Graduate Only)

- A case study of evaluation
- Due 11:59 pm EST, Tuesday, Nov. 14, 2023 in Blackboard

Evaluation Methods

Pre- & Post-prototype

- ✓Surveys: questionnaires
- ✓Surveys: interviews
- ✓Surveys: focus groups
- ✓Functional analysis
- ✓Task analysis

Post-prototype

- ✓Personas
- ✓Cognitive walkthrough
- ✓Card Sorting
- ✓Heuristic evaluation
- ✓Field/ ethnographic
- **User testing**
- **Experiments**

User Testing

User Testing

- Often less “experimental” than normal lab studies
- Used in industry

User Testing - Procedure

- Select a set of tasks (10-30 common tasks)
- Give the user the information required to do the task
- Watch
- Record behaviors of interest

User Testing – In a Lab

Lab environment allows you to capture

- Voice
- Video
- Screen shots
- Keystrokes
- Facial expressions/body language

User Testing Metrics

As your prototype becomes more refined, usability testing becomes more quantitative.

- To collect data, a functional prototype can be built
- Users are given a set of scenarios of tasks that they would perform under usual circumstances

User Testing Metrics

The most common metrics

- Errors
- Time to perform tasks
- Time to perform subtasks
- User subjective reactions (e.g., satisfaction, preferences)

Examples of User Testing Metrics (Detail)

- **Efficiency**
  - Time to learn
  - Time spent on errors
  - Time to complete task
  - Percent or number of errors
  - Frequency of help or documentation use
  - Number of repetition of failed commands

Examples of User Testing Metrics (Detail)

- **Effectiveness**
  - Workload
  - Percent of tasks completed
  - Ratio of successes to failures
  - Number of features or commands used

Examples of User Testing Metrics (Detail)

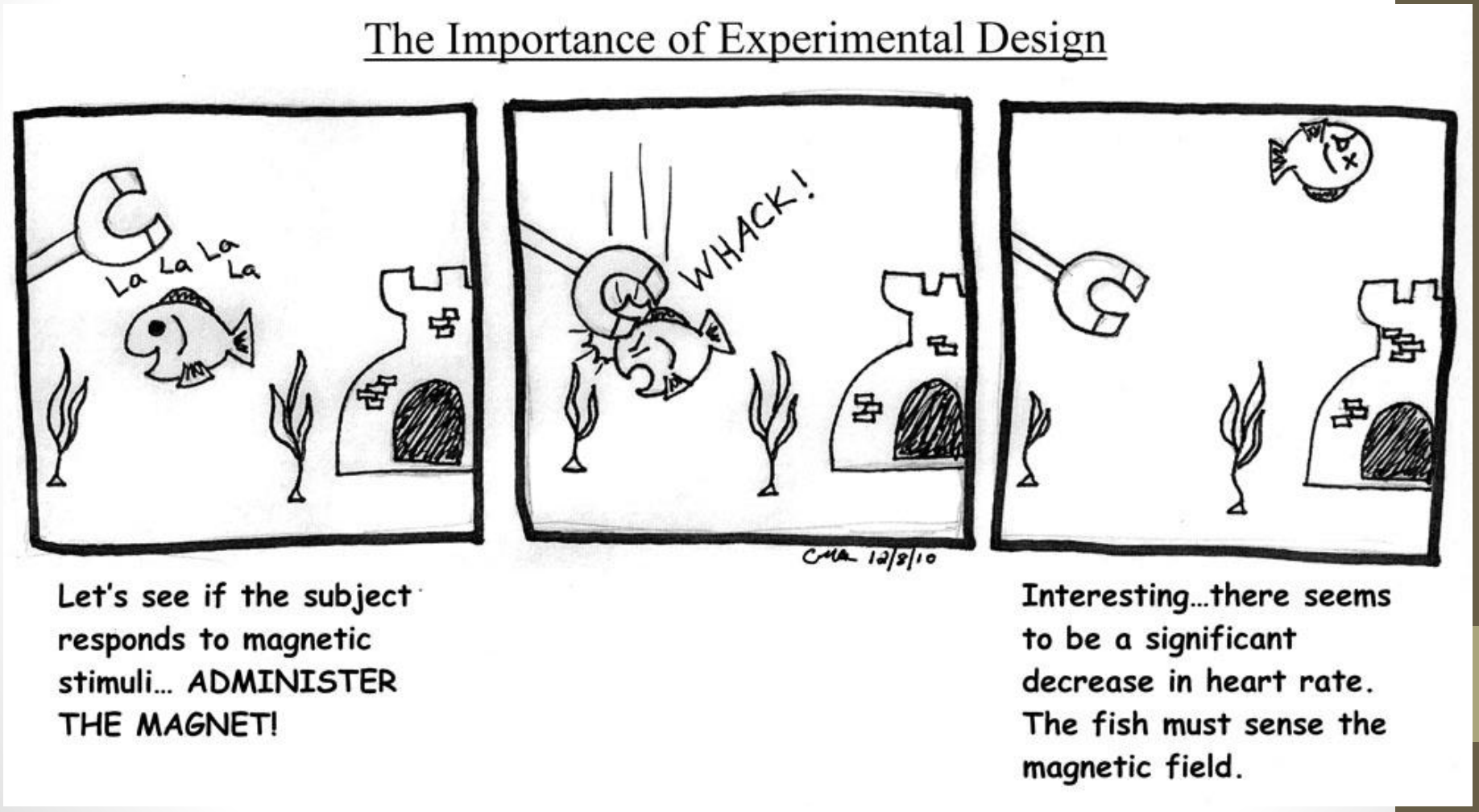
- **User satisfaction**
  - Rating scale for usefulness of software
  - Rating scale for satisfaction with functions/features
  - Perception that the software supports tasks as needed
  - Number of times user expresses frustration or dissatisfaction

Some Suggestions

- **No interference from researcher**
  - Can make user nervous
  - Influence data
- **Listen to the users!**
  - Think aloud protocol, user notes

Experiments

In Your Own Words, What is An Experiment?



Experimental Designs

- Allow the researchers to test specific hypotheses
- Examine relationships between an independent variable and a dependent variable
- Design conditions must be met:
  - Manipulation of an independent variable
  - Measurement of a dependent variable

Independent vs Dependent Variables

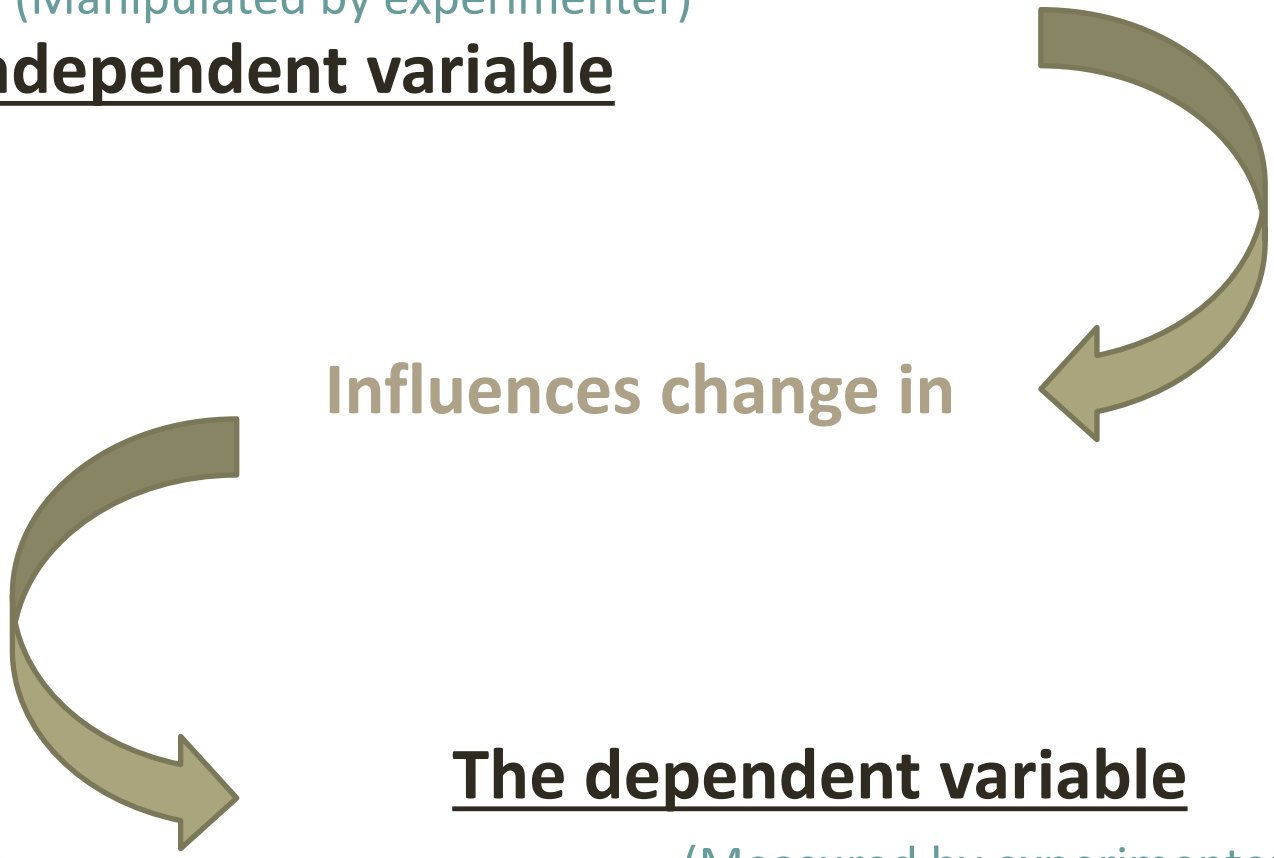
Two types of variables in experimental designs

- The **independent variable** (IV) is *manipulated* by the experimenter
- The **dependent variable** (DV) is *measured* by the experimenter to assess the impact of the independent variable

Independent vs Dependent Variables

(Manipulated by experimenter)

The independent variable



Influences change in

The dependent variable

(Measured by experimenter)

Example

Researchers want to understand how to improve exam grades by manipulating study techniques.

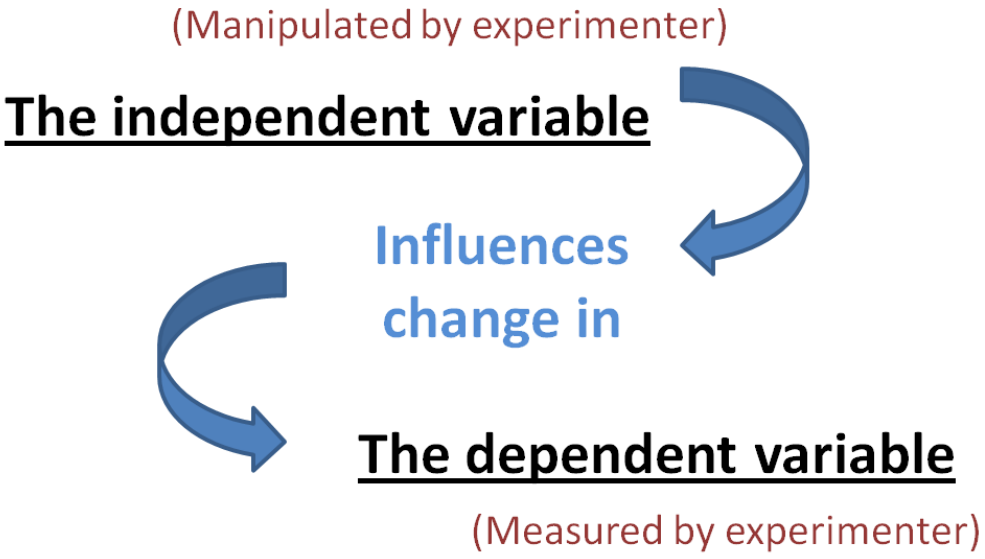
- One group of students study on their own,
- and the other group of students must study with a computerized tutorial (e.g., serious game) periodically throughout their study session.
- Their exam performance is measured.

What is independent variable? What is dependent variable?

Independent variable is study technique. Dependent variable is exam performance.

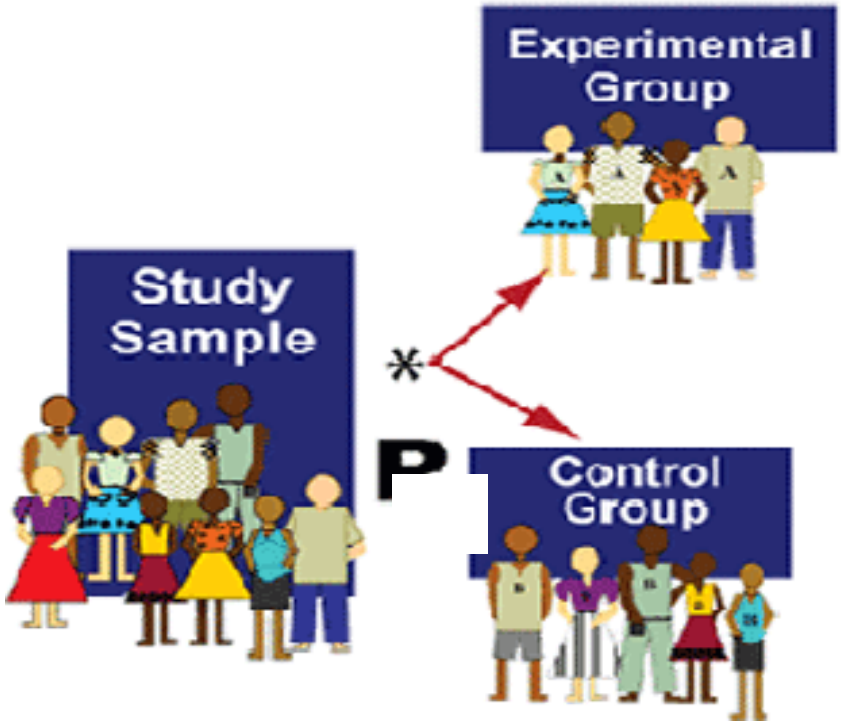
Example

- In the experiment investigating study techniques
- The **independent variable** is study technique
- The **dependent variable** is exam performance



Experimental and Control Groups

- Participants are assigned to one of two groups
  - **Experimental (or Treatment) group** receives some sort of intervention or manipulation
  - **Control (or Comparison) group** receives no intervention



Example

Researchers want to understand how to improve exam grades by manipulating study techniques.

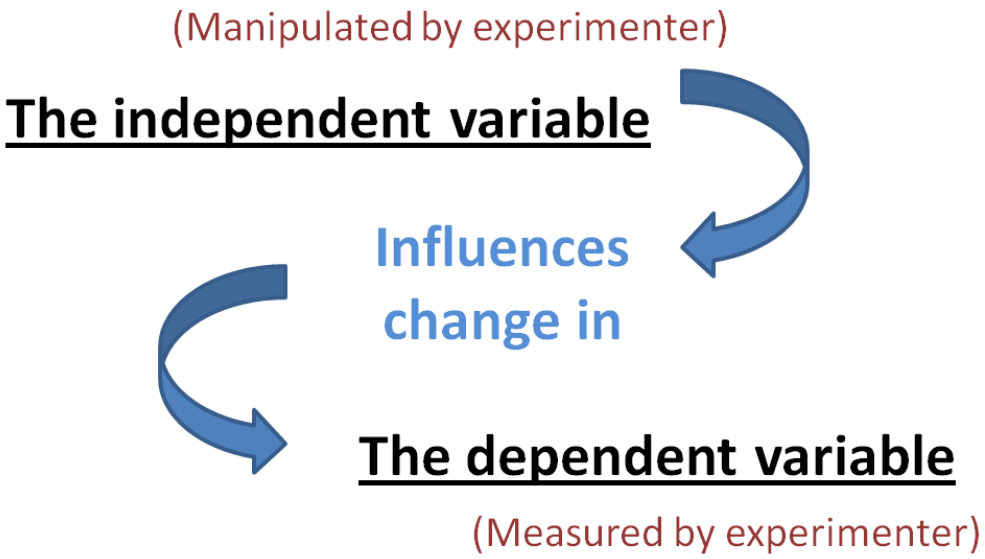
- One group of students study on their own,
- and the other group of students must study with a computerized tutorial (e.g., serious game) periodically throughout their study session.
- Their exam performance is measured.

Which is the control group and which is the experimental group?

The control group is the group of students studying by themselves. The experimental group is the group using the computerized tutorial.

Concept Check 1

A researcher is interested in assessing the impact of alcohol on math test performance. Participants are randomly assigned to either ingest 3 pints of alcohol or 3 pints of water and their performance on the SAT Math subscale is recorded.

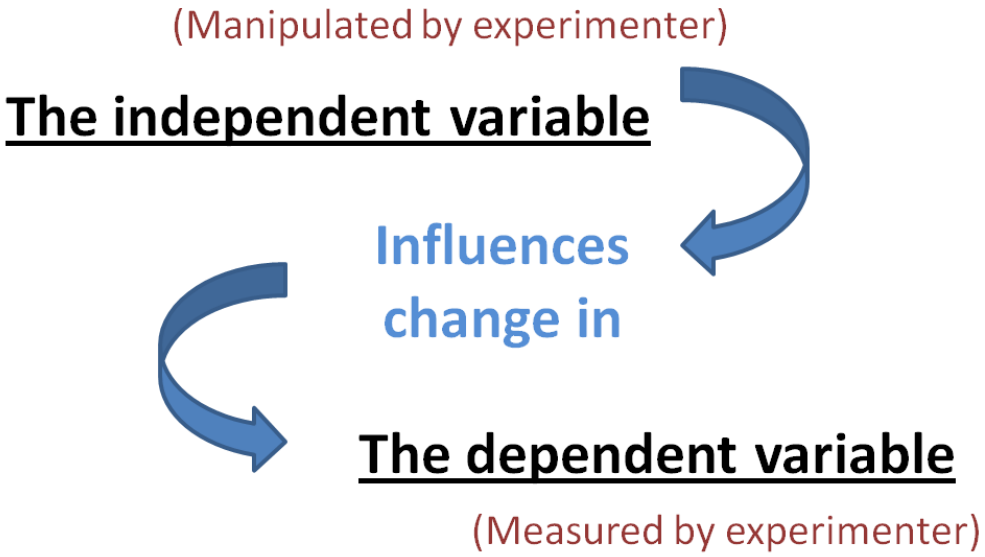


Concept Check 1

- Which is the control group? Which is the experimental group?
  - Control - Alcohol none
  - Experimental – alcohol 3 pints
- What is the independent variable?
  - Alcohol consumption (None vs. 3 pints)
- What is the dependent variable?
  - SAT Math performance

Concept Check 2

A researcher is interested in studying how fatigue influences driving performance. Participants are randomly assigned to either stay up all night or to get normal sleep. Performance in a driving simulator is assessed the next morning.



Concept Check 2

- Which is the control group? Which is the experimental group?
  - Control – normal sleep
  - Experimental – no sleep
- What is the independent variable?
  - Amount of sleep (None vs. Normal)
- What is the dependent variable?
  - Simulated driving performance

Establishing Causality

- A ***confound*** is a variable that affects the dependent variable and may also vary between conditions or groups
  - Provides an alternative explanation for relationship between the independent and dependent variable
  - To establish cause, a researcher must control for all confound variables

Example of Confound

A researcher wants to assess if it is better to space out studying over multiple days or to do all your studying at once. He randomly assigns participants to one of two groups:

- Study 3 hours on Monday
- Study 3 hours on Monday and 3 hours on Tuesday

Both groups take a math test on Wednesday

What is a potential confound? How can we control for this confound, and prevent it from influencing our interpretation?

The potential confound is the studying time.

Another Example of Confound

An educational HCI researcher has developed a new web-based teaching strategy. She uses the new strategy in a gifted class, and the traditional strategy with a regular class.

The effectiveness of the teaching strategy is measured by final course grades.

What is a potential confound? How can we control for this confound, and prevent it from influencing our interpretation?

The potential confound is the difference in knowledge level.

Experimental Designs

- Between-Subjects Design
  - Each participant is assigned to only one of the possible conditions or groups
- Within-Subjects Design
  - Each participant experiences every condition

Concept Check – Between or Within?

20 participants are recruited for a study on the effects of text messaging on driving performance. 10 participants are randomly assigned to perform a simulated drive while responding to text messages, while 10 participants perform a simulated drive without receiving text messages.

Is it a between group or within group design?

- *Between*

Concept Check – Between or Within?

30 doctors are recruited to test the effects of a new stress intervention technique. Each doctor attends an online stress education lecture every day for two weeks, with their self-reported stress evaluated after each week. The next 2 weeks, each of the 30 doctors participate in tai chi, with their self-reported stress evaluated after each week.

Is it a between group or within group design?

- *Within*

STATISTICS

Analyzing Descriptive Data

- ***Descriptive*** statistics
  - Summary of data
  - “describes” the data

Descriptive Statistics

- Measures of Central Tendency
  - **Mode:** Most frequently occurring score
  - **Median:** Halfway point in a set of data
  - **Mean:** Arithmetic average of the scores
- Measures of Variability
  - **Range:** Difference between highest and lowest score
  - **Standard Deviation:** Average difference between each score and the mean of the data set

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

Inferential Statistics

- The possibility that the observed results represents a real and reliable phenomenon
- **Statistically significant:** The likelihood of getting a certain result by chance has a low probability
- HCI typically uses the criterion of $p < .05$
  - Occur by chance less than 5 percent of the time

Ethical Guidelines for HCI Researchers

All researchers must:

1. Obtain informed consent from participants
2. Minimize any discomfort and risk to participant
3. Ensure participants will not suffer any long-term negative consequences
4. Treat any information from participant as confidential
5. Debrief the participant afterward

- The Role of the IRB

What I Expect from You

Nov. 9 – Nov. 14

- Work on prototyping
- Work on the usability test plan
- If you choose a survey, please work on your survey form and send me the form for review.

What I Expect from You

Nov. 16 - Dec. 7

- Perform usability test using the evaluation methods you proposed
- Collect usability testing data
- Analyze usability testing data
- Refine and improve your prototype

Done by Dec. 5

What I expect from you

A report on usability test results is due by 11:59pm EST, Dec. 5

- Title: "CSCE 572 Group X Usability Test Report"
- You can start from the "CSCE 572 Group X Evaluation Methods"
- For each evaluation method you used, describe
  - Rationale
  - How you implemented it, e.g., a procedure
  - The usability data you collected
  - A summary of the data
    - Using metrics if applicable
    - Using descriptive statistics if applicable

What I expect from you

The two deliverables (evaluation methods and test result report) will contribute to your participation grade (10% in your overall grade) and be weighted by peer evaluation

Individual Assignment - Graduate Only

Graduate students are required to participate in usability test:

- Participate in 3 evaluations conducted by other groups
- Apply >= 2 different methods

Individual Assignment - Graduate Only

A report of participation submitted to Blackboard, due: **11:59pm, Nov. 28**

Including

- A summary of evaluation activities you conducted: group ID + evaluation method
- A proof-of-participation for each evaluation activity, e.g.,
  - a snapshot of the card sorting result,
  - a copy/snapshot of completed survey, and
  - a summary of heuristic evaluation with severity rating.
- This assignment will contribute to your participation grade

Extra Credits (Individual)

Participation in usability test

- 1 credit when you participate in one evaluation conducted **by another group**
- Maximum 2 points for the same evaluation method,
  - e.g., you may earn 2 points for 2 surveys, 2 points for 2 heuristic evaluations, and 1 point for 1 user testing
- Maximum 5 points towards your final grade
- **Graduate students need to first meet the requirement of individual assignment before earning extra credit**

Extra Credits (Individual)

To earn the extra credit, you need to submit a participation report to Blackboard, due: 11:59pm, Nov. 28

Including

- A summary of evaluation activities you conducted: group ID + evaluation method
- A proof-of-participation for each evaluation activity, e.g.,
  - a snapshot of the card sorting result,
  - a copy/snapshot of completed survey, and
  - a summary of heuristic evaluation with severity rating.

Teamwork

Work on

- Developing your prototype and
- Developing your evaluation plan