
CS4249 Phenomena and Theories of HCI
Shengdong Zhao
Fall, 2014

Acknowledgement: these slides were adapted from Scott MacKenzie’s course in CHI
What is Empirical Research

• Empirical research is …
  – Observation-based investigation seeking to discover and interpret **facts, theories, or laws**
    • pertinent to phenomena relating to humans interacting with computers
Why do Empirical Research

• We conduct empirical research to …
  – Answer (and raise!) questions about new or existing user interface designs or interaction techniques

  – Develop or explore models that describe or predict behaviour (of humans interacting with computers)
How do we do Empirical Research?

• We conduct empirical research through …
  – A program of inquiry conforming to the *scientific method*
High-level Themes

• Observe and measure
• Research questions
• The 5-step approach
• Statistical analysis
• Report findings
Observe and measure

• Observations are gathered …
  – Manually (human observers)
  – Automatically (computers, software, cameras, sensors, etc.)

• A measurement is a recorded observation

When you cannot **measure**, your knowledge is of a meager and unsatisfactory kind

William Thompson, 1st Baron Kelvin, 1883
Scales of Measurement

- **Nominal**
  - arbitrary assignment of a code to an attribute, e.g.,
  \[ 1 = \text{male}, \ 2 = \text{female} \]

- **Ordinal**
  - rank, e.g.,
  \[ 1^{\text{st}}, 2^{\text{nd}}, 3^{\text{rd}}, \ldots \]

- **Interval**
  - equal distance between units, but no absolute zero point, e.g.,
  \[ 20 \, \text{C}, \ 30 \, \text{C}, \ 40 \, \text{C} \ldots \]

- **Ratio**
  - absolute zero point, therefore ratios are meaningful, e.g.,
  \[ 20 \, \text{wpm}, \ 40 \, \text{wpm}, \ 60 \, \text{wpm} \]

Use ratio measures where possible.
Pop Quiz

• In mathematics, there are three common ways of “averaging”
  – Mode
  – Median
  – Mean (arithmetic)

• Please describe which of the above average method(s) can apply to
  – Nominal data
  – Ordinal data
  – Interval data
  – Ratio data
Pop Quiz

• Which scale of measurement is good for what type of data?
  – Length (?)
  – Opinion (?)
  – Type of gender (?)
  – IQ (?)
  – Temperature (?)
  – Mass (?)
Ratio Measurements

• Preferred scale of measurement
• With ratio measurements summaries and comparisons are strengthened
• Report “counts” as ratios where possible because they facilitate comparisons
• Example – a 10-word phrase was entered in 30 seconds
  – Bad: \( t = 30 \) seconds
  – Good: Entry rate = \( \frac{10}{0.5} = 20 \) wpm
• Example – two errors were committed while entering a 10 word (50 character) phrase
  – Bad: \( n = 2 \) errors
  – Good: Error rate was \( \frac{2}{50} = 0.04 = 4\% \)
High-level Themes

• Observe and measure
• Research questions
• The 5-step approach
• Statistical analysis
• Report findings
Research Questions

• We conduct empirical research to **answer** (and raise!) **questions** about new or existing UI designs or interaction techniques!

• **What are some of these questions?**
  – Below are some examples …
  – Is it viable?
  – Is it as good as or better than current practice?
  – Which of several design alternatives is best?
  – What are its performance limits and capabilities?
  – What are its strengths and weaknesses?
  – Does it work well for novices, for experts?
  – How much practice is required to become proficient?
Are these good questions?
Testable Research Questions

• Preceding questions, while unquestionably relevant, are not testable
• Try to re-cast as testable questions (… even though the new question may appear less important)
• Scenario …
  – You have invented a new text entry technique for mobile phones, and you think it’s pretty good. In fact, you think it is better than the commonly used soft keyboard technique. You decide to undertake a program of empirical
Research Questions (2)

- Is the new technique any good?
- Is the new technique better than soft keyboard?
- Is the new technique faster than soft keyboard?
- Is the new technique faster than soft keyboard within one hour of use?
- If error rates are kept under 2%, is the new technique faster than soft keyboard within one hour of use?
Research Questions (2)

• Is the new technique any good?
• Is the new technique better than soft keyboard?
• Is the new technique faster than soft keyboard?
• Is the new technique faster than soft keyboard within one hour of use?
• If error rates are kept under 2%, is the new technique faster than soft keyboard within one hour of use?
A Tradeoff

Accuracy of Answer

If error rates are kept under 2%, is the new technique faster than soft keyboard within one hour of use?

Is the new technique better than soft keyboard?

Internal validity

Breadth of Question

Breadth of Question

External validity
Internal Validity

• **Definition:** The extent to which the effects observed are due to the test conditions (e.g., soft keyboard vs. new)

• **Statistically …**
  – Differences (in the means) are due to inherent properties of the test conditions
  – Variances are due to participant differences ("pre-dispositions")
  – Other potential sources of variance are controlled or exist equally and randomly across the test conditions
External Validity

• **Definition:** The extent to which results are generalizable to other people and other situations

• Statistically …
  – People
    • The participants are *representative* of the broader intended population of users
  – Situation
    • *Test environment* and *experiment tasks* are representative of real world situations where the interface or technique will be used
Test Environment Example

• Scenario …
  – You wish to compare two interfaces for mobile usage (e.g., text messaging)
• External validity is improved if the test environment mimics expected usage
• Test environment should probably involves…
  – Walking on a real street
  – Let participant to use their own mobile phones
  – Let them type the usual text messages they like to type
• But … is internal validity compromised?
The Tradeoff

- There is tension between internal and external validity.
- The more the test environment and experimental tasks are "relaxed" (to mimic real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as environmental variations, distractions, or secondary tasks.
How can we deal with conflict?
Best of Both Worlds

• Internal and external validity are increased by …
  – Posing multiple narrow (testable) questions that cover the range of outcomes influencing the broader (untestable) questions
    • E.g., a technique that is faster, is more accurate, takes fewer steps, is easy to learn, and is easy to remember, is generally better

• The good news
  – There is usually a positive correlation between the testable and untestable questions
    • I.e., participants generally find a UI better if it is faster, more accurate, takes fewer steps, etc.
As an HCI researcher, your job is to **find** the best set of narrow (**testable**) questions to answer the broader (**untestable**) questions!
One-of vs. Comparative

- Many “user studies” in HCI are one-of
  - I.e., a new user interface is designed and a user study is conducted to find strengths and weaknesses
- Much better to do a comparative evaluation
  - I.e., A new user interface is compared with an alternative design to determine which is better
- The alternative may be
  - A variation in the new design
  - An established design (perhaps a “baseline condition”)
- More than two interfaces may be compared
- Testable research questions are comparative!
- See the paper in CHI 2006 by Tohidi et al.
High-level Themes

- Observe and measure
- Research questions
- The 5-step approach
- Statistical analysis
- Report findings
The 5-Step Approach to Design Controlled Experiments
Outline

The 5 Step Approach to Experiment Design

1. Define the research question
2. Determine variables
3. Arrange conditions
4. Decide blocks and trials
5. Set instruction and procedures
Let’s Start with an Example

Problem

earPod vs. iPod
Menu Selection on Mobile Devices Often Requires Visual Feedback
Why Eyes-free?
Why eyes-free?
Visual vs. auditory menu

Visual Linear Menu

IVR System
earPod
earPod design

Dial

Inner disc
Novice Usage
Intermediate I
Intermediate II
Expert Usage
Video

- [http://www.youtube.com/watch?v=bATkA0Usoio](http://www.youtube.com/watch?v=bATkA0Usoio)

Paper

Question: earPod vs. iPod
The 5 Step Approach to Experiment Design

1. Define the research question
2. Determine variables
3. Arrange conditions
4. Decide blocks and trials
5. Set instruction and procedures
The 5 Step Approach to Experiment Design

1. **Define the research question**
   - Step 1.1 Start with a general question and change it to a specific one
   - Step 1.2 Define target population
   - Step 1.3 Define task(s)
   - Step 1.4 Define measure(s)
   - Step 1.5 Define factor(s)

2. Determine variables
3. Arrange conditions
4. Decide blocks and trials
5. Set instruction and procedures
Step 1.1: Start with a General Question

How does earPod compare with iPod’s menu in terms of performance?
Step 1.1: Specific Research Question

- How does your technique compare with other alternative technique(s)
  - Techniques
- for what tasks
  - Tasks
- in what context
  - Other factors
- in terms of what kind of measures
  - Performance measures
- for which target population
  - Target users
Step 1.1: Specific Research Question

- How does your technique \(\rightarrow\) earPod
- compare with other alternative technique(s) \(\rightarrow\) iPod
  - Techniques
- for what tasks
  - Tasks
- in what context
  - Other factors
- in terms of what kind of measures
  - Performance measures
- for which target population
  - Target users
Step 1.2: Define Target Population

General question: How does earPod compare with iPod’s menu in terms of performance?

Target population?
Question: we designed earPod for whom?
• How does your technique compare with other alternative technique(s) – Techniques
  – Tasks
• for what tasks – Tasks
• in what context – Other factors
• in terms of what kind of measures – Performance measures
• for which target population – Target users
Step 1.3: Define Task(s)

**General question:** How does *earPod* compare with iPod’s menu in terms of performance?

**Task(s):** menu selection

However, the menu selection task has endless possibilities: single short menu, single long menu, hierarchical menus.
Step 1.3: Define Task(s)

Key insight: experiment design need to decide what subset of tasks is appropriate to test.

Question: how do you choose the subset?
• How does your technique earPod
• compare with other alternative technique(s) iPod
  – Techniques
• for what tasks
  – Tasks
• in what context
  – Other factors
  – Performance measures
• in terms of what kind of measures
• for which target population
  – Target users
Step 1.4: Define Measures

**Question:** How does earPod compare with iPod’s menu in terms of performance?

**Measures:** performance

In HCI, we typically use three measures to quantify performance:
- Speed
- Accuracy
- Learnability

**Key insight:** need to define “testable” measures
• How does your technique compare with other alternative technique(s) – earPod
  – Techniques
• for what tasks – Tasks
• in what context – Other factors
• in terms of what kind of measures – Performance measures
• for which target population – Target users
Step 1.5: Define (other) Factors

**Question:** How does earPod compare with iPod’s menu in terms of performance?

**Factors:** other than the different type of tasks, what other factors can influence the measures?

**Again:** the number of factors are unlimited …

- Scenario of use
- Input device
- Background of the user
  - Educational level
  - Gender
  - Ethnic background
  - Age
  - …

**Key insight:** experiment design need to determine a subset of factors to test.

**Question:** how to choose the factors?
Let’s Review Step 1

Step 1: Define the research question
  – Step 1.1: Start with a general question
  – Step 1.2: Define target population
  – Step 1.3: Define task(s)
  – Step 1.4: Define measure(s)
  – Step 1.5 Define factor(s)
Let’s Practice

Example 1: “earPod vs. iPod”

1.1: General Question
   – How does earPod compare with iPod’s menu in terms of performance?

1.2: Target Population
   – Young generation

1.3: Task(s)?
   – e.g., Menu selection for three types of breadth (4, 8, 12) and two types of depth (1, 2), content of the menu is from common categories

1.4: Measures?
   – Speed, accuracy, learning

1.5: (Other) Factors?
   – Single-task vs. multi-tasking
   – …
Example 2: “Opti” vs. “Qwerty” Keyboard

1.1: General question
   – how does the “opti” keyboard layout compare with the “qwerty” keyboard in performance?

1.2: Target Population?
   – Computer users?

1.3: Task(s)?
   – Type “the quick brown fox jumps over the lazy dog”

1.4: Measure(s)?
   – Speed, accuracy, learning

1.5: (Other) Factors?
   – Device: Touch typing vs. stylus?
   – Screen size: different screen size?
The 5 Step Approach to Experiment Design

1. Define the research question
2. **Determine variables**
3. Arrange conditions
4. Decide blocks and repetitions
5. Set instruction and trials
Step 2: Define Variables

Type of variables

• Independent variable (IV)
  – Factors that are manipulated in the experiment
  – Have multiple levels

• Dependent variable (DV)
  – Factors which are measured

• Control variable
  – Attributes that will be fixed throughout experiment
  – Confound – attribute that varied and was not accounted for
    • Problem: Confound rather than IV could have caused change in DVs
  – Confounds make it difficult/impossible to draw conclusions

• Random variable
  – Attributes that are randomly sampled
  – Increases generalizability
Type of Independent Variables

• Primary
  – The most important independent variable(s) that you want to investigate

• Secondary
  – The other interesting factors you want to manipulate in the experiment.
  – They help to answer the main question in a richer way.
Task Type & Factor  ➔ Independent variables

Measures  ➔ Dependent variables

Everything else  ➔ Control/Random Vars.
Let’s Try

Example 1: earPod vs. iPod

• **Independent variables**
  - Technique
    • 2 levels (earPod vs. iPod)
  - usage scenario
    • 2 levels (single-task vs. dual-task)
  - menu breadth
    • 3 levels (4, 8, 12)
  - menu depth
    • 2 levels (1, 2)

• **Dependent variables**
  - Speed (measured in completion time)
  - Accuracy (measured in percentage of errors)
  - Learning (measured in speed & accuracy change over time)
Let’s Try

Example 1: earPod vs. iPod

• Control variables
  – Same computer, experiment time, environment, instruction, etc.

• Random variables
  – Attributes of participants: age, gender, background, etc.
Let’s Try Again

Example 2: “Opti” vs. “Qwerty” keyboard layout

- **Independent variables**
  - Type of keyboard
    - 2 levels (opti vs. qwerty)
  - Input method
    - 2 levels (touch vs. stylus)
  - Screen size
    - 3 levels (watch, mobile phone, tablet)

- **Dependent variables**
  - Speed (measured in word per minute)
  - Accuracy (measured in?)
  - Learning (measured in speed & accuracy change over time)
Example 2

Example 2: “Opti” vs. “Qwerty” keyboard layout

• Control variables
  – Same computer, experiment time, environment, instruction, etc.

• Random variables
  – Attributes of participants: age, gender, background, etc.
Confounding Variable
Pepsi vs. Coke

http://www.youtube.com/watch?v=ebSMyWuO7eE
Which One Tastes Better?

M  PEPSI

Coca-Cola

Q
Which One Tastes Better?

- M (PEPSI)
- Q (Coca-Cola)

- M (Coca-Cola)
- Q (PEPSI)
Testing Result:

65% of people like Pepsi (the drink in M cup)

**Conclusion:** People like Pepsi’s taste better than Coke

**Question:** Do you trust this result?
Which One Tastes Better?

M

PEPSI

Coca-Cola

Q
Which One Tastes Better?

M | M
---|---
PEPSI | Coca-Cola

Q | Q
---|---
Coca-Cola | PEPSI
Testing Result:

65% of people like Coke (the drink in M cup)
What have we learned?

• Design a fair experiment is tricky
• There are many factors that can influence the result
  – Some we want to control and understand
  – Some we want to eliminate (confound)
Confounding Variable

Any variable other than the independent variables that can possibly explain the change in measures

Example 1 – two techniques are compared (earPod, iPod)
• All participants are tested on earPod, followed by iPod
  – Performance might improve due to practice
  – The order of presenting the technique is a confounding variable (because it explains the changes in measures but it is not an IV)

Example 2 – two software interfaces are compared (Microsoft Word vs. new)
• All participants have prior experience with Microsoft Word, but no experience with the new interface
  – “Prior experience” is a confounding variable

Note: Order of presentation & Prior experience are two important confounding variables we need to control. More on this later …
Review

• What are the 4 different scales of measurement?
  – What’s their difference?
  – Which scale of measure is preferred? Why?

• What do you think of the following research question? Any issues?
  – Is my technique viable?

• What is internal validity? What is external validity? What’s the relationship between these two?

• How to convert a rough research question into a testable research question?

• What are the 4 different types of variables for experiment design?
The 5 Step Approach to Experiment Design

1. Define the research question
2. Determine variables
3. **Arrange conditions**
   From Independent Variables to Experimental Conditions
4. Decide blocks and trials
5. Set instruction and procedures
What is a condition?

• Let’s start with an example

• A particular independent variable “Technique” has two levels: earPod and iPod.
  – If it is the only independent variable considered, this experiment has two conditions

• However, an experiment rarely only has 1 independent variable, suppose there is another independent variable “Menu Breadth” with 3 levels (4, 8, 12).
  – There are 2 (Techniques) x 3 (Menu Breadth) = 6 experimental conditions
  – The each unique combination of the different levels of the various independent variables (such as earPod, 4) is an experimental condition
• earPod
  – Breadth 4
  – Breadth 8
  – Breadth 12

• iPod
  – Breadth 4
  – Breadth 8
  – Breadth 12
How can We Test These Conditions?

• **Method 1:**
  – Recruit 6 participants, one for each condition (this is also called *between-subject design*, which means the conditions are tested between different subjects)
    • P1: earPod, 4
    • P2: earPod, 8
    • P3: earPod, 12
    • P4: iPod, 4
    • P5: iPod, 8
    • P6: iPod 12
  – What’s the problem of this approach?
    • What about individual differences?
    • To balance individual differences, we need lots of participants

• **Key insight:** this method is expensive
How can We Test It?

• **Method 2:**
  - Recruit the same participants to test all 6 conditions (this is also called *within-subject design* since all conditions are tested within the same subject)
  - This method is much more economical
  - What’s the problem of this approach?
    • Practice (or order effect) as a confounding variable
    • However, in many cases, this effect can be controlled
Control Order Effect using Counter-balancing

If we assume the order effect is symmetric, which means $A \rightarrow B = B \rightarrow A$, and is linear, which means the increment between different conditions is about the same, we can use **counter-balancing** to cancel the effect out.

E.g., we assume the transferring effect between (A after B) and (B after A) are both 10

Participant 1: A followed by B (A B)
Participant 2: B followed by A (B A)

**Observation**: the order effect equally affects both A and B, so the absolute relationship between A and B is not changed.

**However**, a minimum number of participants is needed for counter-balancing to work
Counter-balancing with 3 Levels

What if an IV has 3 levels? A B C

If the same assumption holds: assume effects are symmetric, and equal in size. We need to counter-balance as follows.

P1: A B C
P2: A C B
P3: B A C
P4: B C A
P5: C A B
P6: C B A

What about 4 levels, 5 levels, 6 levels, …?
4 level = 4! (24), 5 levels = 5! (120), …
Introducing Partial Counter-balancing: Latin Square

Latin square:
- Ensures each level appears in every position in order equally often:

\[
\begin{align*}
A & \quad B & \quad C \\
B & \quad C & \quad A \\
C & \quad A & \quad B
\end{align*}
\]

Assume \( A-B = B-A = A-C = C-A = B-C = C-B = 10 \)

\[
\begin{align*}
P1: \quad & a + (b+10) + (c+20) \\
P2: \quad & b + (c+10) + (a+20) \\
P3: \quad & c + (a+10) + (b+20)
\end{align*}
\]

Average \( A = (3a +30)/3 = a + 10 \)
Average \( B = (3b+30)/3 = b + 10 \)
Average \( C = (3c+30)/3 = c + 10 \)

However, \( A-B = B-A = 10 \)
\( A-C = C-A = 20 \)
\( B-C = C-B = 30 \)

\[
\begin{align*}
P1: \quad & a + (b+10) + (c+50) \\
P2: \quad & b + (c+30) + (a+30) \\
P3: \quad & c + (a+20) + (b+40)
\end{align*}
\]

Average \( A = (3a +50)/3 = a + 50/3 \)
Average \( B = (3b+50)/3 = b + 50/3 \)
Average \( C = (3c+80)/3 = c + 80/3 \)
Steps for Arranging Conditions for Within-Subject Design

3.1: List all Independent Variables and their levels
3.2: Decide counter-balancing strategy for each variable
3.3: Determine the minimum No. of participants
3.4: Arrange the overall design
3.5: Determine detailed arrangement for each participant
Example 1: earPod vs. iPod

Assume we have three IVs

Step 3.1: list the IV and their levels
- Technique (2 levels: earPod, iPod)
- Scenario of use (2 levels: single-task, multi-task)
- Menu depth (2 levels: 1, 2)

Step 3.2: determine counter-balancing strategies for each IV
- Choices: 1) fully counter-balancing, 2) Latin-square, 3) no counter-balancing (sequential)
- Question: how to decide which strategy to use?
  - It depends on how interesting is the independent variable
  - It depends on how much resource we have
Example 1: earPod vs. iPod

Step 3.2: Counter-balancing strategies
- Technique (2 levels, fully counter-balance)
- Scenario of use (2 levels, fully counter-balance)
- Menu depth (2 levels, no counter-balance, sequential)
- Why?

Step 3.3: Determine the minimum No. of participants
- Minimum No. = 2 Tech. conditions X 2 Scenario conditions X 1 Menu depth arrangement = 4
- Question: if Menu depth is also fully counter-balanced, how many participants we need?
- Question: if Technique has 3 levels and it is fully counterbalanced (assume menu depth is not counter-balanced), how many participants we need?
Step 3.4: Determine the overall arrangement

T1, T2  x  Single-task, Multi-task
T2, T1  Multi-task, Single-task

Condition arrangement

Step 3.5: Determine arrangement for each participant
In-class Exercise: Example 2

Step 3.1: List IVs
- Technique (3 levels: A, B, C)
- Scenario of use (2 levels: single-task, multi-task)

Step 3.2: Decide counter-balancing strategy

Step 3.3: Determine Minimum No. of Participants

Step 3.4: Determine the overall arrangement

Step 3.5: Determine individual arrangement for each participant
However, Counter-balancing May not Always Work

Counter-balancing Assumes symmetric transfer and linear increment
  - A-B transfer == B-A transfer
  - A-B transfer == B-C transfer

If asymmetric transfer and non-linear increment
  - i.e., A-B transfer > or < B-A transfer or A-B <> B-C
  - Have to use **Between-subjects design**
  - In addition, some factors have to be between-subject
    - Age, Gender, etc.
No. of Condition Reduction Strategies

In experiment design, one major problem we often face is there are many possible relevant factors. It’s important for experiment designers to pick the most important/interesting factors to test.

Run a few independent variables at a time
  – If strong effect, include variable in future studies
  – Otherwise pick fixed control value for it

Not all within-subject IVs need to be counter-balanced
  – If we are not interested in the absolute difference among different levels, we don’t need to counter-balance. E.g., Menu Breadth, Menu Depth, etc.
Exercise: earPod vs. iPod

• Independent variables
  – Technique (2 levels: earPod vs. iPod)
  – Usage scenario (2 levels: single vs multi-tasking)
  – Menu breadth (3 levels: 4, 8, 12)
  – Menu depth (2 levels: 1, 2)

• Question: which of these factors need counter-balancing?
Let’s Review: Between- vs. Within-Subject Design

• Method 1: use a lot of participants, randomly assign them to each technique (between-subject design)
  – Drawback: costly

• Method 2: use the same participant to test both techniques (within-subject design)
  – Drawback: practice effect
For Between-subject Design, there is no need for counter-balancing, just assign different users to different conditions!
Steps for Arranging Conditions for Within-Subject Design

3.1: List all Independent Variables and their levels
3.2: Decide counter-balancing strategy for each variable
3.3: Determine the minimum No. of participants
3.4: Arrange the overall design
3.5: Determine detailed arrangement for each participant
The 5 Step Approach to Experiment Design

1. Define the research question
2. Determine variables
3. Arrange conditions
4. **Decide blocks and trials**
5. Set instruction and procedures
Definitions

• **Trial**
  – A single repetition of a single condition/cell
  – Each trial in a condition is treated as equivalent with each other
  – Typically you want to have at least 3 trials per condition to increase reliability
  – The number of trials is determined by the sample space

• **Block***
  – An entire section of the experiment
  – Repeated to analyze learning

* Block has other definitions. This is a simplified definition for the purpose of this assignment.
Block: same arrangement, repeated

Trials in each block: same content, with order randomized
An example

• Condition: iPod, Breadth 4
  – Trial: select menu item 1
  – Trial: select menu item 2
  – Trial: select menu item 3
  – Trial: select menu item 4

• P1 (iPod, Breadth 4)
  – Item 3
  – Item 1
  – Item 2
  – Item 4

• P2 (iPod, Breadth 4)
  – Item 1
  – Item 4
  – Item 3
  – Item 2
Block Indicates Learning

Adapted from the TiltText paper by Widgor & Balakrishnan
Determine Number of Blocks/Repetitions

• Reasonable experiment duration
  – Time Constraint and Fatigue
  – Typically within 1 hour
    • However, minus pre- and post-experiment interviews, only left with 45 minutes
  – In some cases, up to 2 hours

• Enough data points for significant effects
Step 4: Determine Blocks and Trials

- Step 4.1: estimate the time for each trial (typically at least 3 trials per condition)
- Step 4.2: estimate the time for each block
- Step 4.3: balance the trials and blocks so that the main part of the experiment is within 45 minutes
- Step 4.4: combine with the condition arrangement
Exercise

Full experiment design: earPod vs. iPod

Independent variables

- Technique (2 levels: earPod vs. iPod)
- Usage scenario (2 levels: single vs multi-tasking)
- Menu breadth (3 levels: 4, 8, 12)
- Menu depth (2 levels: 1, 2)

Block = 3
Trials per condition = 4
Each trial takes roughly 10 seconds to finish

Question: how is the experiment arranged?
Question: how long will the experiment take?