Warm up

Researchers would like to investigate whether being read to by teachers improves children’s reading comprehension. Fifty children in an inner-city elementary school were tested during an after-school program for children of working parents. All children were first given an identical reading comprehension pre-test. Half of the children in the study were read to three times a week by their teacher during the after-school program, and half of the children spent extra time with their teacher three times a week during the after-school program doing other activities. At the end of the three week period, all of the children were given an identical reading comprehension post-test that differed from the pre-test they taken at the beginning of the study. This research project potentially has:

(a) Low External Validity
(b) Instrumentation threat
(c) Low Construct Validity
(d) History Threat

Pragmatic Issues

- Nice job with posts
  – Any questions?
- Any questions about Assignment 1?
- Assignment 2 will ask you to perform a part of Cognitive Task Analysis in an educational subject-matter domain of your interest
  – What’s your off-the-cuff idea for that domain?

Experiments

- How is an experiment a distinct type of study?
- How do experiments differ from other types of empirical research?
**Things all experiments have in common**

- Two or more conditions, with everything identical except for the factor(s) being tested
- Some form of randomization in assignment of subjects to conditions

**Experiments**

- What types of validity are (generally) higher with experiments?
  - Internal
  - Construct
  - External
  - Ecological
  - Conclusion

**Experiments**

- Are any types of validity generally **lower** with experiments?
  - Internal
  - Construct
  - External
  - Ecological
  - Conclusion
Causal Conclusions

• The big thing: it is easier to infer causality in experiments than in other types of studies

• (It’s not impossible to do otherwise, see Richard Scheines’s lectures)

• That said…

Internal Validity in Experiments

• Is confounding a bigger problem in educational research or, say, medicine?

Confounding

• One excellent way to mess up an experiment is *confounding*

• When you vary two or more things at a time

• Confounding is a violation of what kind of validity?

Ways Experiments Can Vary

• Number of conditions
• Simple conditions or crossed factors
• Pre-test or not
• Randomized assignment or Blocked randomized assignment
• Within-subjects, between-subjects, or crossover
• Laboratory/artificial setting or classroom/natural setting
## Ways Experiments Can Vary

- Number of conditions
- Simple conditions or crossed factors
- Pre-test or not
- Randomized assignment or Blocked randomized assignment
- Within-subjects, between-subjects, or crossover
- Laboratory/artificial setting or classroom/natural setting

## Number of conditions

- Classical experiment: One control condition, one experimental condition

---

## Number of conditions

- Three condition experiment:
  - One control condition, two experimental conditions
  - Two control conditions, one experimental condition

## Number of conditions

- Three condition experiment:
  - One control condition, two experimental conditions
  - **Two control conditions, one experimental condition**

- Why might you want to do this?
Ways Experiments Can Vary

- Number of conditions
- **Simple conditions or crossed factors**
- Pre-test or not
- Randomized assignment or Blocked randomized assignment
- Within-subjects, between-subjects, or crossover
- Laboratory/artificial setting or classroom/natural setting

Crossed factors

- Two by two “factorial” design:
  - Two factors that are either control or experimental

---

**Two by two experiment:**

*Example* (Mathan & Koedinger, 2002)

<table>
<thead>
<tr>
<th>Expert</th>
<th>Feedback</th>
<th>Intelligent Novice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Walkthroughs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO Example Walkthroughs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Why would you want to do this?**

<table>
<thead>
<tr>
<th>Expert</th>
<th>Feedback</th>
<th>Intelligent Novice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example Walkthroughs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO Example Walkthroughs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Benefits to factorial design

- Investigate more factors in a single study
- Allows you to investigate interaction effects
  - Does Intelligent Novice feedback only work when the student already has sufficient conceptual knowledge (from the example walkthrough)?

Important to remember

- You need a much bigger sample size to be able to find statistical significance for an interaction effect (to put it another way, interaction effects have lower statistical power)

More complex factorial designs

- 3x3
- 2x2x2
- 5x4x7x8x2

- Why don’t we see more 5x4x7x8x2 factorial designs in published papers?

Ways Experiments Can Vary

- Number of conditions
- Simple conditions or crossed factors
- Pre-test or not
- Randomized assignment or Blocked randomized assignment
- Within-subjects, between-subjects, or crossover
- Laboratory/artificial setting or classroom/natural setting
Do you agree?

- Trochim and Donnelly suggest that pre-tests are not necessary

- Do you agree?

Post-test only comparison

When you only test post-test, you assume...

But what if

This condition is better!

They were significantly different at pre-test!

Happens 5% of time by chance
But what if

They were significantly different at pre-test!
Happens 5% of time by chance

PRE POST

Another problem with post-test only designs

Pretty bad control condition, isn’t it?

PRE POST

Can you defend Trochim?

• Trochim says that this is not a problem
• “Even if you find that the groups differ on a pretest, you have no reason to suspect that they will differ on a posttest. Why? Because their pretest difference had to be a chance one.”

• How might this make sense?

Can you guarantee group equivalence?

• “just because you randomly assign people to groups does not guarantee those groups are equivalent and does not guarantee the treatment effect you find is not due to differences in groups.”
• In fact, there is no way in general to guarantee equivalent groups
  – You can measure things of interest, but
    1) Measures have error
    2) You might not be measuring the right things, that is, even if your groups are equivalent on your measures, like pre-test, they might non-equivalent on something you did not measure, like IQ, which may matter
Pre-tests can help

• How?

• What downsides?

Ways Experiments Can Vary

• Number of conditions
• Simple conditions or crossed factors
• Pre-test or not
• Randomized assignment or Blocked randomized assignment
• Within-subjects, between-subjects, or crossover
• Laboratory/artificial setting or classroom/natural setting
Randomized Assignment

• Easy idea

• Every student is *randomly* assigned to a condition

Is this randomized assignment?

• Aleven
• Baker
• Corbett
• Dabbish

• Evenson
• Forlizzi
• Hudson
• Koedinger

CONTROL condition

EXPERIMENTAL condition

Why is this worse than randomized assignment?

• Aleven
• Baker
• Corbett
• Dabbish

• Evenson
• Forlizzi
• Hudson
• Koedinger

Blocked Randomized Assignment

• You do this when you think sub-groups of your population will be different from each other

• Within each group, assign subjects randomly to conditions
For example

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTROL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ways Experiments Can Vary

- Number of conditions
- Simple conditions or crossed factors
- Pre-test or not
- Randomized assignment or Blocked randomized assignment
- Within-subjects, between-subjects, or crossover
- Laboratory/artificial setting or classroom/natural setting

Between-subjects

- Each student is in either the control or experimental condition

CONTROL

EXPERIMENTAL

Within-subjects

- Every student is in both the control and experimental conditions

CONTROL

EXPERIMENTAL
Crossover

- Every student is in both the control and experimental conditions, in both possible orders

```
CONTROL     EXPERIMENTAL
CONTROL     EXPERIMENTAL
```

What needs to be true, for each approach to be valid?

- Between-subjects
- Within-subjects
- Crossover

What are some dangers for each approach?

- Between-subjects
- Within-subjects
- Crossover
What are some dangers for each approach?

• Between-subjects
  – Ethical concerns, compensatory rivalry, resentful demoralization

• Within-subjects
  – What if there is some learning between conditions?

• Crossover
  – What if something happens during the study, making two halves of study incommensurate?
  – Also twice as much work

Ways Experiments Can Vary

• Number of conditions
• Simple conditions or crossed factors
• Pre-test or not
• Randomized assignment or Blocked randomized assignment
• Within-subjects, between-subjects, or crossover
  • Laboratory/artificial setting or classroom/natural setting

What are some

• Benefits of lab studies

What are some

• Costs of lab studies
What are some

- Benefits of field studies

What are some

- Costs of field studies

Types of Validity

- Lab Studies

- Field Studies

Validity

Closing Thoughts

- There are a lot of ways that experiments can vary from one another
Ways Experiments Can Vary

• Number of conditions
• Simple conditions or crossed factors
• Pre-test or not
• Randomized assignment or Blocked randomized assignment
• Within-subjects, between-subjects, or crossover
• Laboratory/artificial setting or classroom/natural setting

But…

• There are some characteristics all experiments share
  – Two or more conditions, with everything identical except for the factor(s) being tested
  – Some form of randomization in assignment of subjects to conditions

• Studies that don’t have these characteristics still may be fine studies, they’re just not experiments