The Ten Big Ones
(for Experimental Research in Accounting)

Kristy L. Towry
Emory University
ABO Doctoral Colloquium 2014

Why Rules?

0 A set of guidelines for good practice.
0 Think of them like grammar.
0 Great scholars sometimes break them.
0 Following them won’t make you great.
  0 As a young researcher, it might keep you from making “rookie mistakes.”
Caveats

- If we talk about commandments, we must talk about sinners.
- Some examples are just made up.
- For others, elements are disguised to protect the (not) innocent.
  - Sometimes not very effectively.
- Too many managerial examples.
  - Not because we have more sinners.

The Background

CONSTRUCTS

PROXIES
Danger! Danger!

The First Big One

Thou shalt think at the construct level.
Nobody cares what happened in your laboratory except you.
It’s difficult, because experimentalists love our instruments.
This means you should write your hypotheses at the construct level!
The First Big One
Example

- What is the effect of “peer monitoring” on honesty in self-reported performance?
  - My friend’s theory: “If my peers are watching, I’ll report more honestly, because I think my peers will think badly of me if I lie.”
    - Peer monitoring will make me more honest.
  - My theory: “If my peers are watching, I’ll report less honestly, because my peers will think badly of me if my performance is too low.”
    - Peer monitoring will make me less honest.

The First Big One

Danger! Danger!
If you can’t explain your theory without describing your instrument, you’re not thinking at the construct level.
The Second Big One

Thou shalt not make conclusions about effect sizes based on the results of an experiment.

Why?

Effect sizes are somewhat arbitrary in the laboratory.
They depend on parameter choices.

The Second Big One

Example

I want to see what has the biggest effect on ethical decision making:

A mission statement focusing on ethics, or
An explicit policy for penalizing ethical breaches.
Study Design

- 2 X 2 between-subjects design.
  - IV1: Mission statement: Y or N
  - IV2: Penalty policy: Y or N
  - DV: Explicit choice
A Debate for Experimentalists

> Economic significance is more important than statistical significance.
> Research needs to address economic significance.
> Does this admonition apply to experimentalists?

The Third Big One

> Thou shalt not conduct a horse race unless the two horses (i.e., theories) are mutually exclusive.
Thou shalt not test for an interaction by comparing p-values.

First, what is an interaction?
The effect of one variable depends on the level of another.

Example
I predict that the effect of audit expertise on task performance depends on task complexity.

Formally stated hypothesis:
The effect of audit expertise will be more positive for more complex tasks.
Predictions / Results

Results:

• When the task is complex, the effect of expertise is significantly positive \((p < 0.01)\).
• When the task is simple, expertise has no effect \((p > 0.05)\).

Has the theory been supported?

Would You Believe?

Results:

• When the task is complex, the effect of expertise is significantly positive \((p < 0.01)\).
• When the task is simple, expertise has no effect \((p > 0.50)\).

Has the theory been supported?
**In Other Words**

*The only way to test for an interaction is to test for an interaction.*

**The Fifth Big One**

*Thou shalt not test for an interaction (in a 2 X 2 design) by seeing if the mean in one condition is different from the mean of the other three conditions.*
The Fifth Big One Example

I am interested in understanding how “tone at the top” affects how investors judge the credibility of management disclosures.

Theory: It depends on whether the disclosure is consistent or inconsistent with expectations.

Predictions

Formally Stated Hypothesis:

- When there is a weak ethical tone and the disclosure is inconsistent with expectations, investors will judge credibility lower than in any other condition.
**Predictions / Results**

- Results: Credibility judgment is lower in the inconsistent / weak condition than in the
  - inconsistent / strong (p < 0.05),
  - consistent / strong (p < 0.05), and
  - consistent / weak (p < 0.05) conditions.

**Would You Believe?**

- Results: Credibility judgment is lower in the inconsistent / weak condition than in the
  - inconsistent / strong (p < 0.05),
  - consistent / strong (p < 0.05), and
  - consistent / weak (p < 0.05) conditions.
In Other Words

- The only way to test for an interaction is to test for an interaction.

The Sixth Big One

- Thou shalt not worry about mundane realism.
  - A “realistic looking” experiment is not the goal.
    - (What is the goal?)
  - “Simulating the real world is not the goal.
    - (We already have the real world.)
The Sixth Big One

- You will be asked about your design choice.
- Don’t answer: I did it like that, because it’s how things are done in the real world.
- Do answer: I did it like that, because it led to a more effective test of my theory.

The Sixth Big One
Example

- This one’s personal.
- I wanted to study the effects of “partial incentive contracts” on effort toward rewarded vs. unrewarded tasks.
- As an aside, why would you ever use partial incentive contracts?
- You can’t put incentives on all tasks, because some are difficult to measure.
Experimental Design

- Two-task setting
- 2 X 1
- IV: Flat Wage vs. Partial Incentives
- DVs: Performance on Rewarded / Unrewarded Tasks

Picking the Two Tasks

- Remember the “aside”?
  - Why would you want a partial incentive contract?
    - Differentially Measurable Tasks
- Do we want our tasks to be differentially measureable?
  - (It would increase mundane realism)
  - How would it affect our ability to test our theory?
The Seventh Big One

Thou shalt not generalize from one condition, only from comparisons across conditions.

Example

I want to study various factors that affect auditors’ success in identifying financial statement errors.

My instrument includes 10 “seeded errors.”

Result: Across all conditions, auditors identified an average of 3.2 errors.

Conclusion: Auditors are useless.
The Eighth Big One

0Thou shalt not call a study an experiment, unless it involves a manipulation.

The Eighth Big One

0Just because you used a laboratory doesn't mean you ran an experiment.
0Just because people read a case study doesn't mean you ran an experiment.
0Just because people performed a computer task and you measured performance doesn't mean you ran an experiment.
The Eighth Big One

(Is it just semantics?
(What is the experimentalist’s comparative advantage?

The Ninth Big One

(Thou shalt not describe predictions in terms of conditions, but rather, in terms of effects.
(This is especially true when the conditions reflect “relative positioning.”
The Ninth Big One
Example

Based loosely on my colleague Vic Anand’s dissertation. (I’ve revised his hypotheses to make my point better. Don’t judge Vic.)

Hypothesis 1: Shorter goal horizons increase effort when uncertainty about the effort-output relationship is low.

Hypothesis 2: Shorter goal horizons decrease effort when uncertainty about the effort-output relationship is high.

A Better Way

Predict the interaction!

The effect of goal horizon depends on uncertainty, such that as uncertainty increases, the effect of a short goal horizon becomes less positive (or more negative).
The Tenth Big One

\textit{Thou shalt not describe the operationalizations / proxies before describing the constructs.}

The Tenth Big One Example

\textit{Before} the hypotheses are stated:

Our setting is a modification of the capital budgeting setting used in a number of recent experimental studies... In our setting, the superior has a portfolio of projects from which to choose and sufficient means to fund all projects in the portfolio... Each project is proposed by a different subordinate, and the actual cost associated with each of the projects is stochastic... The actual surplus of each project is greater than or equal to zero for all possible realizations of the cost... Revenue and the probability distribution over costs are common knowledge; however, only the subordinate ever knows the actual cost...
Your Turn

Any Confessions?