



Assessment Design

NYU
Assessment
Bootcamp

Where are we at?



What's Assessment Design and Why Worry About It?



- An assessment design is concerned with the who, what, when, where, and how of an assessment.
- The design is driven by the why.
- It is only with a good design that we can find the **themes**, **patterns**, **trends**, and **relationships** we're looking for in our assessments

Here's what we're going to look at...



- **Why** we conduct assessments the way we do
- **How** we will collect our data, and **what** method we will use
- **Who** will participate in our assessments (and the issues related to the “who”)
- **When** the assessments will take place

“Why”

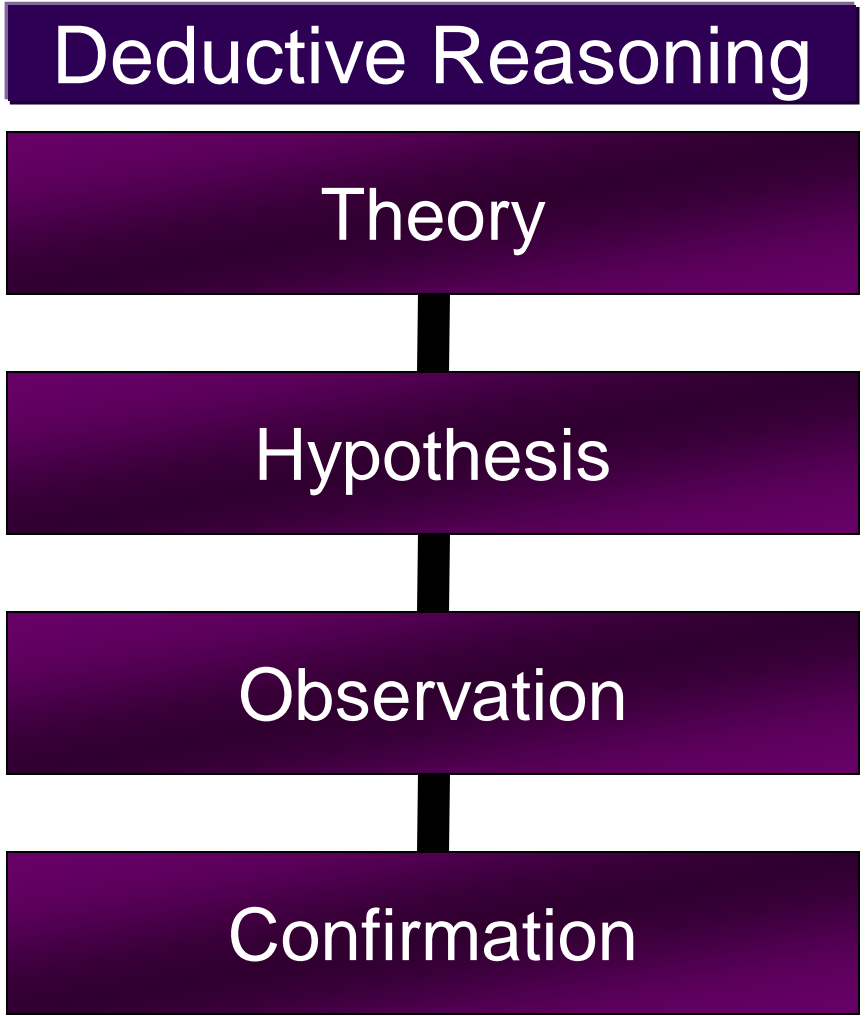
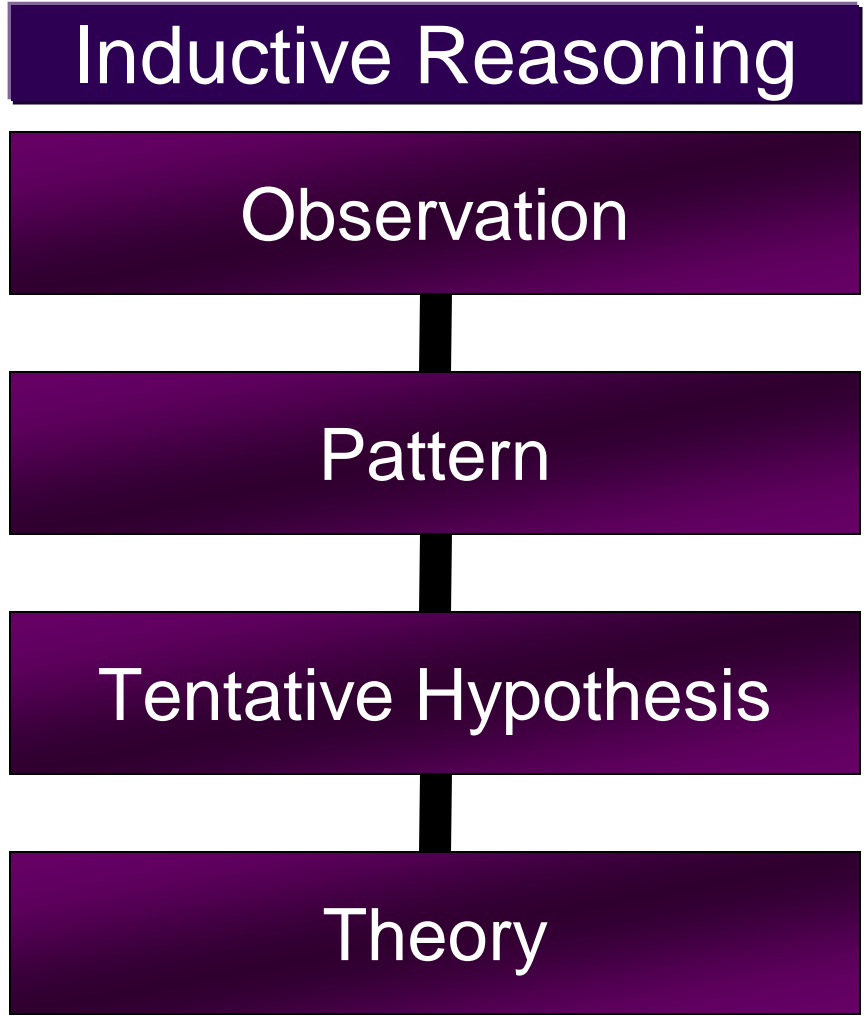


NEW YORK UNIVERSITY

*As assessment practitioners,
we phrase our questions in
terms of whether or not a
given program's outcomes
have been met.*

Created by Peter Swerdewski -
Copyright 2007

“How” & “What [method]”



Qualitative!

Created by Peter Swerdevsk
Copyright 2007

Quantitative!



- In identifying who will take an assessment, the goal is to ensure the results from the assessment are as **valid** and **generalizable** as possible
 - If possible, assess *everyone!*
 - Only interpret your results in terms of who was assessed
 - For qualitative assessment, you want “saturation”
 - For quantitative assessment, you want representativeness
 - Randomly assign students to the program and control group, if possible. If not, try to find a relatively similar control group

How do we know we're making the right conclusions about our people?

Threats to the validity of your results! ☹️

- **Selection**

Students tend to self-select into programs and into assessments

- **History**

Any other event that occurs between pretest and posttest that the groups experience differently



- **Maturation:**
Results from differential rates of normal growth between pretest and posttest for the groups
- **Pre-testing:**
When there is a differential effect between groups on the posttest because of taking the pretest (“priming”)
- **“Mortality”:**
Arises when there is *differential* nonrandom dropout between pretest and posttest



A few words on selection threats:

- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!
- Convenience sampling is not acceptable!

...just don't do it if you can help it!!



Single Post-Test Design

- Easy and frequently-used 😊
- Lacks a pretest 😞
- There is no comparison group, so it's tough to know if the change occurred due to the program or to something else. 😞

Program → Test

Common for:

- Programs offered only once
- Programs for which a big assessment isn't worth the energy



Program → Test

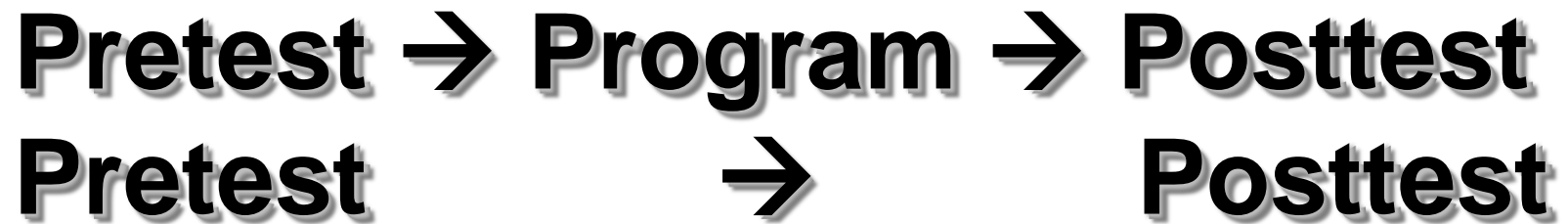
Program → Test

Program → Test

Cross-Sectional Design

- Easy design 😊
- Allows for an understanding of change-over-time 😊
- No pre-test 😞
- Must be careful to interpret results in terms of (a) the attendees and (b) the specific program offered 😞

- Useful for comparing semester-to-semester or year-to-year programs:
 - How well do diversity program satisfaction scores fare from year to year?
 - If we change a residence hall program during year three, will posttest score increase (thus the change was good)?



Pretest / Posttest Design with a Control Group

- Can show **change** *due to the program* 😊😊😊
- *Note:* You have another group that is not participating in the program
(the control group)

- Learning communities
- Learning resource centers
- Career prep courses

An Orientation Program design

- Three waves of administrations to help identify *when* a change actually takes place 😊
- Lots of measuring 😞

Pretest over the summer before students arrive to NYU

Summer orientation

Test just after summer orientation

Welcome Week

Test just after Welcome Week

Freshman year

Test at the end of freshman year

How do we know we're making the right conclusions?



Careful not to believe the wrong thing:

- Believing there's **no relationship** when **there is one**:
 - Needle in a haystack / noise phenomenon
 - Low reliability
 - Random “sameness” (heterogeneity) of respondents
 - Low statistical power
- Believing there's **a relationship** when there is **not one**:
 - “Fishing” in the data for something
 - Capitalizing on chance
 - HARKing: **H**ypothesizing **A**fter **R**esults **K**nown



1. In assessment, we frame our questions in terms of how well we achieve our _____?
2. In assessment, we often act like _____ trying to build _____ for the effectiveness of our program.
3. Why is assessment design important?
4. Why does Pete hate convenience sampling?
5. What's so good about pretests?
6. ***BONUS***: What would a strong assessment design for the Assessment Bootcamp look like?



A few words about assessment designs:

- Assessments are all about **compromise**.
- Often it's helpful to start with an **ideal design**, then work backward to what's plausible.
- It's up to **you** to decide what evidence you will accept as sufficient to answer your assessment questions
- Always **interpret** your results in light of your design.



- We determine our data collection when we determine our assessment design
- Think through how the assessment will be administered
- Minimize motivation issues
- Standardizing → control!
- You don't need to measure every objective every year!
- Again, an assessment is a series of **COMPROMISES!**



This presentation heavily borrows from these excellent resources:

Trochim, William M. The research methods knowledge base, 2nd Ed. Internet WWW page, at URL: <<http://trochim.human.cornell.edu/kb/index.htm>.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). Experimental and quasi-experimental designs. Houghton Mifflin Company: Boston.