What will it take to do effectiveness research?

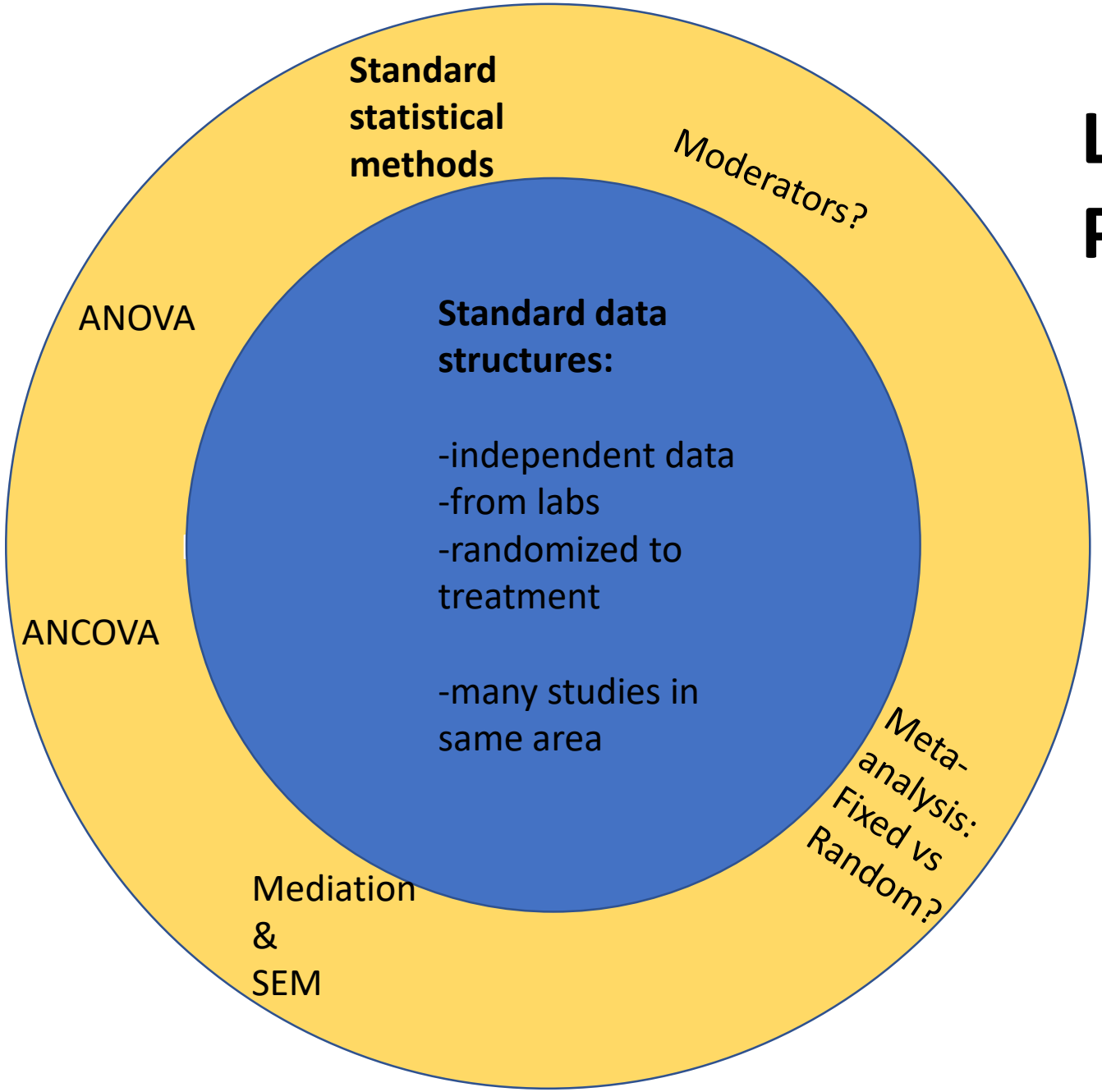
Elizabeth Tipton

Teachers College, Columbia University

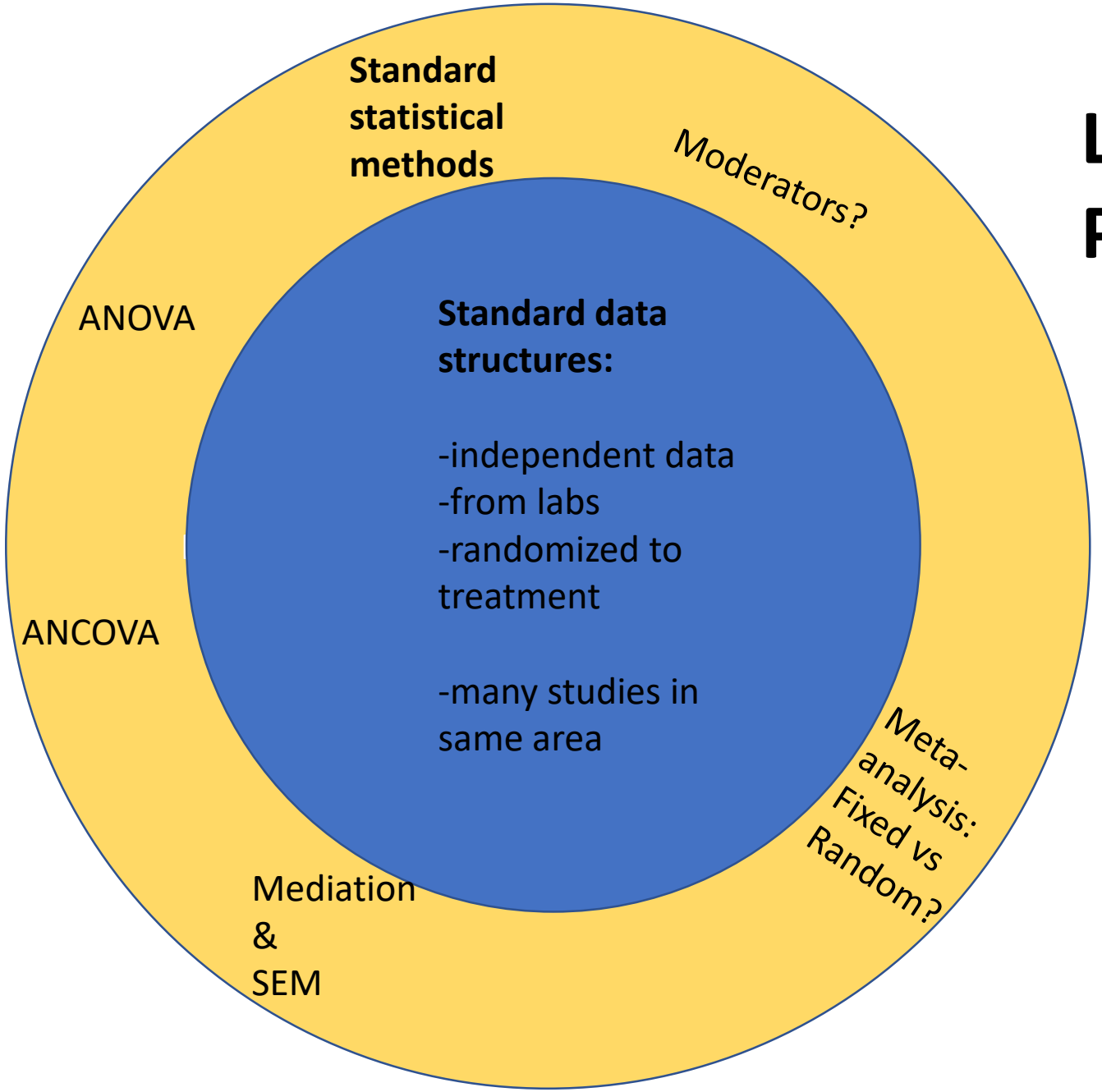
May 25, 2017

Laboratory Psychology

Laboratory Psychology

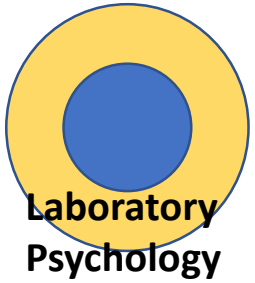


Laboratory Psychology

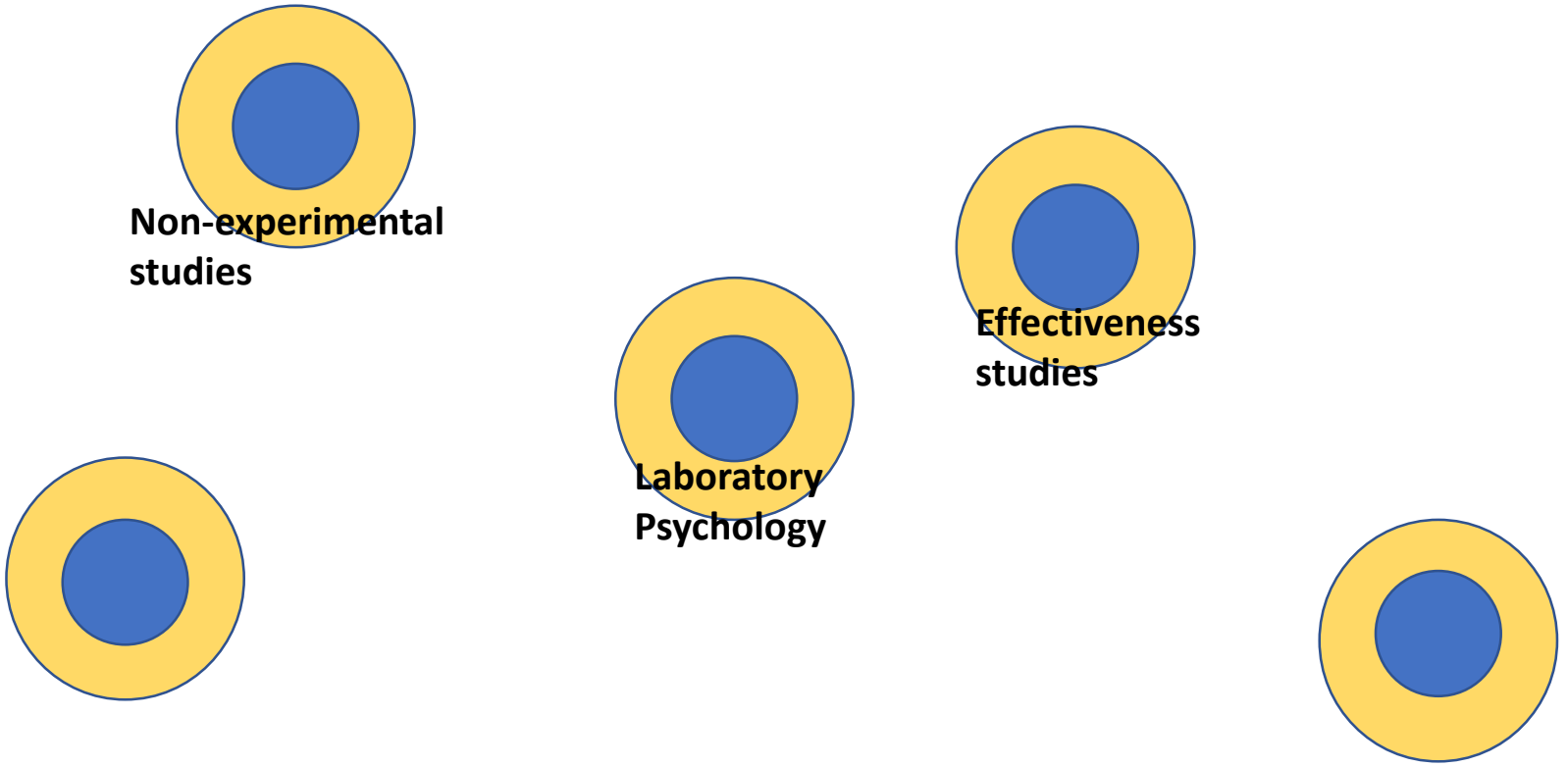


Reminder:
These methods are appropriate for *these data structures*.

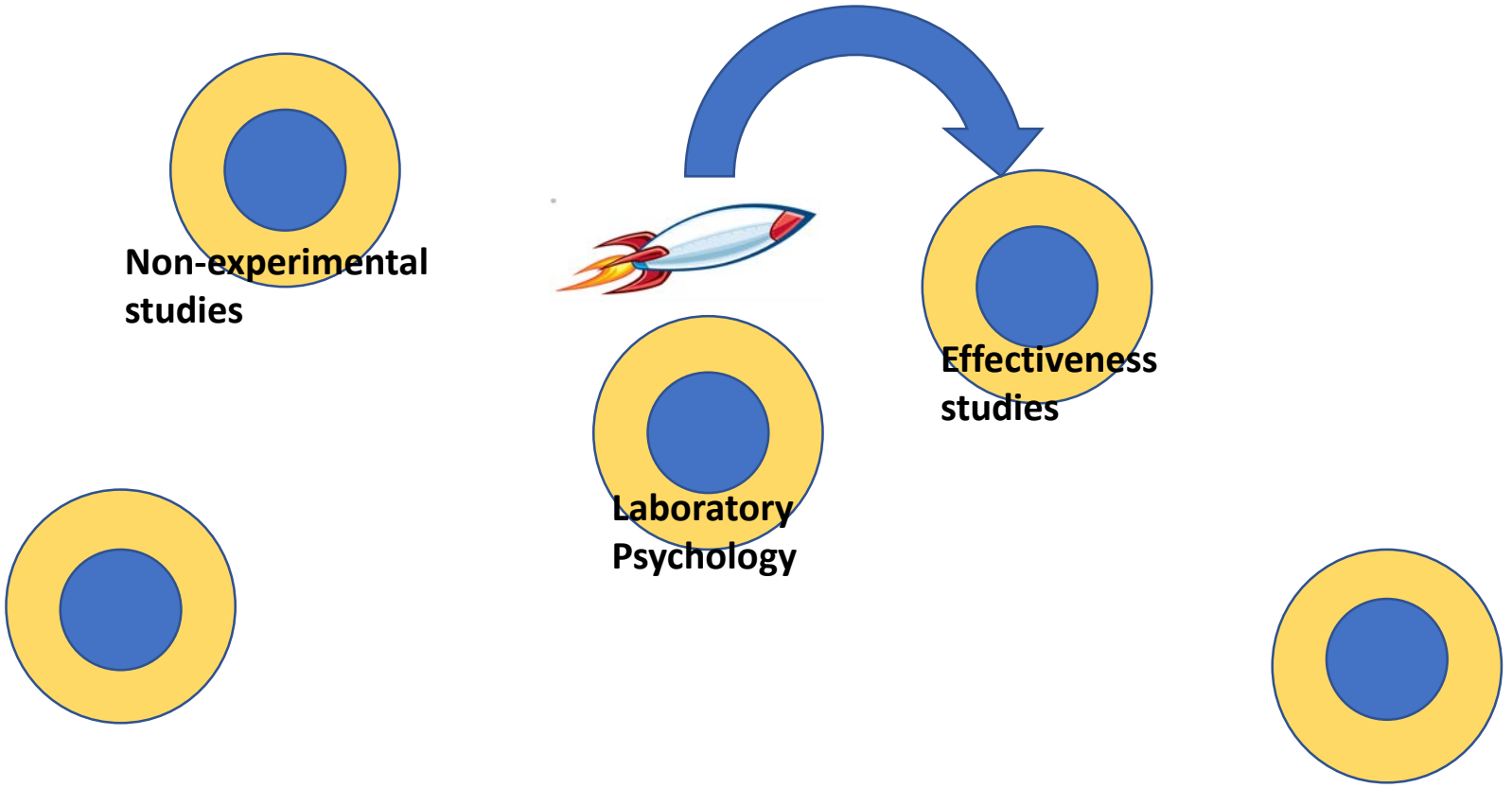
But what happens when we -> effectiveness?



But what happens when we -> effectiveness?



But what happens when we -> effectiveness?



New experimental protocols

Context changes: Lab -> Classrooms/Schools

<u>And thus changes to:</u>

- Recruitment : Students -> Teachers/Schools
- Interventions : Decontextualized -> Contextualized
- Comparisons : Carefully designed -> Business as usual
- Moderators : College age -> K – 12 age
  Baseline knowledge -> Lower/more variable

- Implementation : Researcher -> Teacher/Student
- Outcomes: Researcher developed -> School interested

New data structures

Laboratory studies	Effectiveness studies
Independent units	Nested units (students < teachers < schools)
Individual random assignment	Cluster random assignment
Focus on a single ATE	Possibility of treatment effect variation / interactions
How does it work? (Mechanism)	Will it work in practice? Where?

#1 Random assignment changes

- In laboratories - > individual random assignment
- In schools, it is typically hard to randomly assign individual students
  - How to implement both T/C in the same class or school?
  - What if those in T share with those in C?
  - Sometimes principals or teachers want all students in the same condition.

- This means you may need to recruit schools or teachers, then randomize schools, teachers, or classrooms to conditions

Nested data structures

- When groups are randomized, we have to take this into account in the design and analysis.
  - If students are randomly assigned to these classrooms/teachers/schools, this isn't a problem.
  - But when they are not, students in the same classrooms/teachers/schools are more similar to one another -- > "clustering".
- Designs:
  - Cluster randomized trial (CRT): the level of recruitment and assignment are the same (e.g., teachers recruited and teachers RA)
  - Random block design (RBD)/ multi-site trial (MST): the level of assignment below the level of recruitment (e.g., schools recruited, teachers RA)

Nesting affects analyses

- Now ANOVA isn't appropriate: errors are correlated.
- HLM is a generalization to ANOVA.
- Now power is affected not just by the number of subjects (n) but also by:
  - The number of clusters (m)
  - The *intra-class correlation* (ICC)

- The number of clusters randomized is the most important here.
  - A study with 10 kids in each of 10 randomly assigned classes has *lower* power than one with 100 kids randomly assigned on their own.

#2 Implementation issues

- How would this intervention be implemented in a classroom?
  - Practical concerns: does it require technology? Is this technology available in classrooms?
  - Does it require teacher knowledge?
  - How will this fit into a school day?
  - What is business as usual?

- Don’t assume – test and measure!
  - An amazing intervention that only works when implemented perfectly isn’t realistic. Your intervention needs to be robust and work within the classroom context.
  - How can you measure:
    - How well it is implemented in the treatment group?
    - How well it is implemented in the control group? (Maybe they are getting something quite similar in the business as usual condition)

Implementation analyses

- Treatment is randomly assigned. Implementation of treatment is not.
- In the effectiveness language, we have two estimates:
  - ITT: Intent to Treat Effect -> the policy question
  - TOT: Treatment on the Treated Effect -> the scientific question

- Analyses of implementation have to be careful of confounding.
  - There are methods – e.g., instrumental variables, Bloom's correction
  - These are different than mediation analyses
  - There is new work on how to tease apart causality in mediation

#3 Generalization concerns

- In lab studies, the focus is on mechanism. The population under study is largely based on convenience.
- In effectiveness studies, we want to understand if it works in a **population.**

- If treatment effects are constant, it doesn't matter which schools or students are in our study.
- If treatment effects vary then the ATE *depends on the sample*.

Generalization approaches

- You need to understand schools and context. Where might an intervention work? (Inclusion/Exclusion criteria)
- A single study doesn't have to focus on a broad population. It could focus on a more narrow population and question.

- How?
  - Define an inference population.
  - Recruit the variety of students/teachers/schools found in the population.
  - Compare the types of students/teachers/schools in your study to those found in the population.

- Also: think more carefully about variation
  - Treatment effect moderators?

Example: National Study of Learning Mindsets

- Intervention:
  - Computer delivered
  - Student randomization
  - Brief (< 2 hours) delivered over 2 separate times within 1 semester
- Outcomes:
  - Mindset related
  - Administrative data
- Population:
  - “Regular” public high schools (9-12 grades) in the U.S.

NSLM Study design

- Contracted with a survey research firm
- Randomly selected 140 high schools throughout the US
  - 76 high schools agreed to take part
- Within each high school, all 9th graders were taken to a computer lab and randomized to the intervention within the software

- Study design and goals:
  - Estimate ATE, subgroup ATEs, and variation in TEs across schools
  - Test hypotheses about treatment effect moderators

Conclusions

But I'm not doing effectiveness trials yet …

- You don't have to **move** to the new planet yet. You can **visit** and begin **building capacity.**

- Getting from here to effectiveness is gradual. You'll need to learn about:
  - School contexts
  - Group randomized designs
  - New statistical methods

- But you don't have to recreate the wheel. Other parts of psychology and education have done this:
  - The NSLS suggests that with computer delivered interventions, effectiveness and scalability are easier and are doable.
  - There are training workshops at conferences and over the summer.
  - The WWC guidelines provide an overview.

Resources

- 2 week workshop on CRTs, IES funded, at Northwestern
- 1 week workshop on advanced meta-analysis, at Loyola Chicago

- Optimal Design software (power analysis)
- The Generalizer web tool : www.thegeneralizer.org

- Society for Research on Educational Effectiveness
  - Workshops, some online videos
  - Journal
  - Conference