Mediation
Intro to Analysis & Design

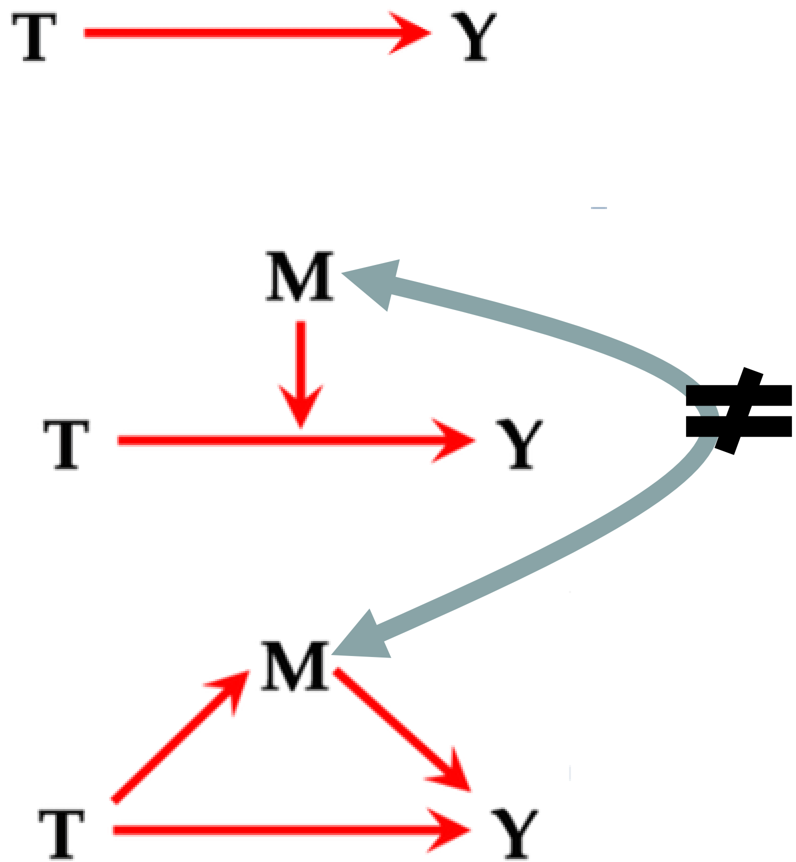
Ben Kelcey
University of Cincinnati

Session

- Introduction to mediation
- Analysis of mediation studies
- Design of mediation studies

Main, Moderation, Mediation

- Main
- Moderation
- Mediation

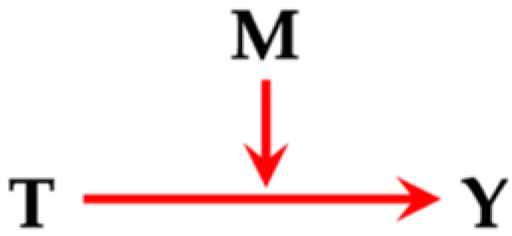


Main Effects

$$T \longrightarrow Y$$

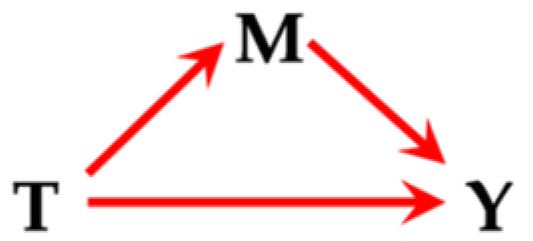
- Most studies initially focus on main or average effects
  - Main effects describe whether a program works on average
- Main effects studies are limited descriptions of a program

Moderation



- Do effects vary by subgroup or context?
- A moderating variable is a pretreatment variable that interacts with the treatment such that the impact of the treatment depends on the value of the moderator variable
- Moderation addresses for whom and under what circumstances a treatment is effective

Mediation



- Mediation analyses unpack the pathways or intermediate variables through which the treatment operates on the outcome
  - Tests a theory of action
  - Informs how a treatment works and identifies any breaks in a theory

Moderation v. Mediation

- Moderation: heterogeneity in treatment effects across subgroups or conditions/contexts
  - Introduces interaction between treatment and pretreatment variables to probe differential effects
- Mediation: probes the mechanisms through which a treatment impacts an outcome
  - Introduces intermediate variables (post-treatment but pre-outcome) that lie on the causal pathway from the treatment to outcome to probe the theory of action

Example Context

A seismic shift in national research priorities over the past 6 years has led to a dramatic increase in the number of large-scale randomized experiments designed to test the impact of educational interventions on student outcomes. Spybrook (2007) identified 55 such trials supported by the Institute for Education Sciences. Of these, the vast majority assigned groups, typically schools or classrooms rather than individuals, to interventions. The majority of the innovative interventions attempted to improve student learning by improving classroom teaching.

Raudenbush, S., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. Journal of Research on Educational Effectiveness, 1(2), 138–154.

Context

A major aim of these studies is to evaluate the impact on student learning of assignment to an innovative classroom intervention. This aim can be achieved, in principle, without measuring the quality of classroom instruction. However, the interpretation of findings from such a study will typically be ambiguous.

Context

Consider a study in which the assignment of schools or classrooms to a novel instructional innovation is found to have no significant impact on student learning. Assume that the study design was unbiased and provided adequate statistical power to detect a nonnegligible effect. Two explanations immediately arise. Program evaluators refer to these as "theory failure" versus "implementation failure" (Rossi, Lipsey, & Freeman, 2004).

Context

First, it may be that the innovation changed classroom instruction in the ways intended but that those classroom changes made no difference in student learning. The term *theory failure* describes this scenario because the theory that links intended changes in instruction to intended student outcomes will have proven incorrect.

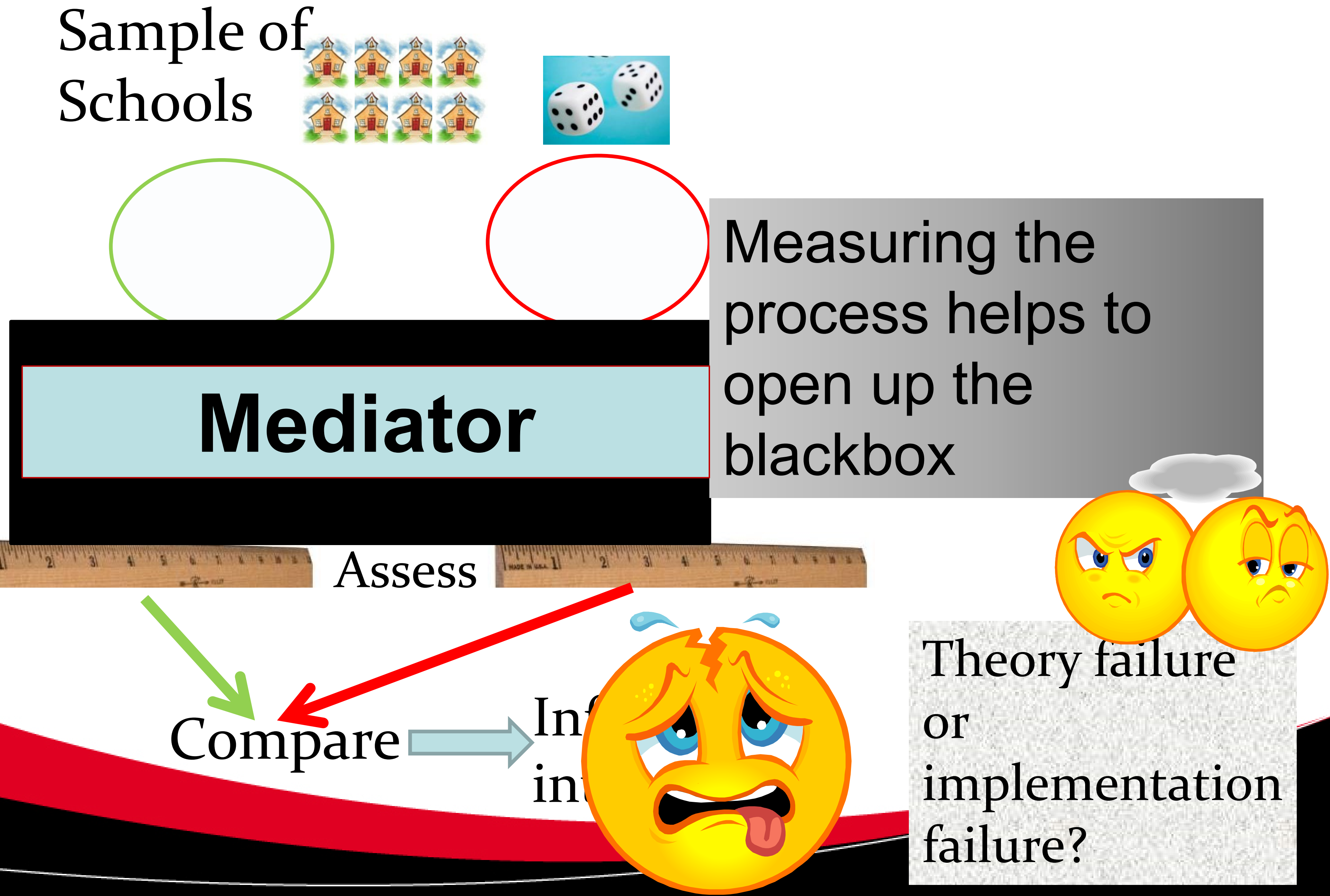
Second, the innovation may never have been effectively implemented in classrooms. Perhaps the innovators lacked skill in working with teachers or perhaps the teachers lacked the skill, knowledge, or motivation to put the innovative ideas to work in their teaching. In any case, program theory about the relationship between the intended instruction and student outcomes was never tested, leading to "implementation failure."

Context

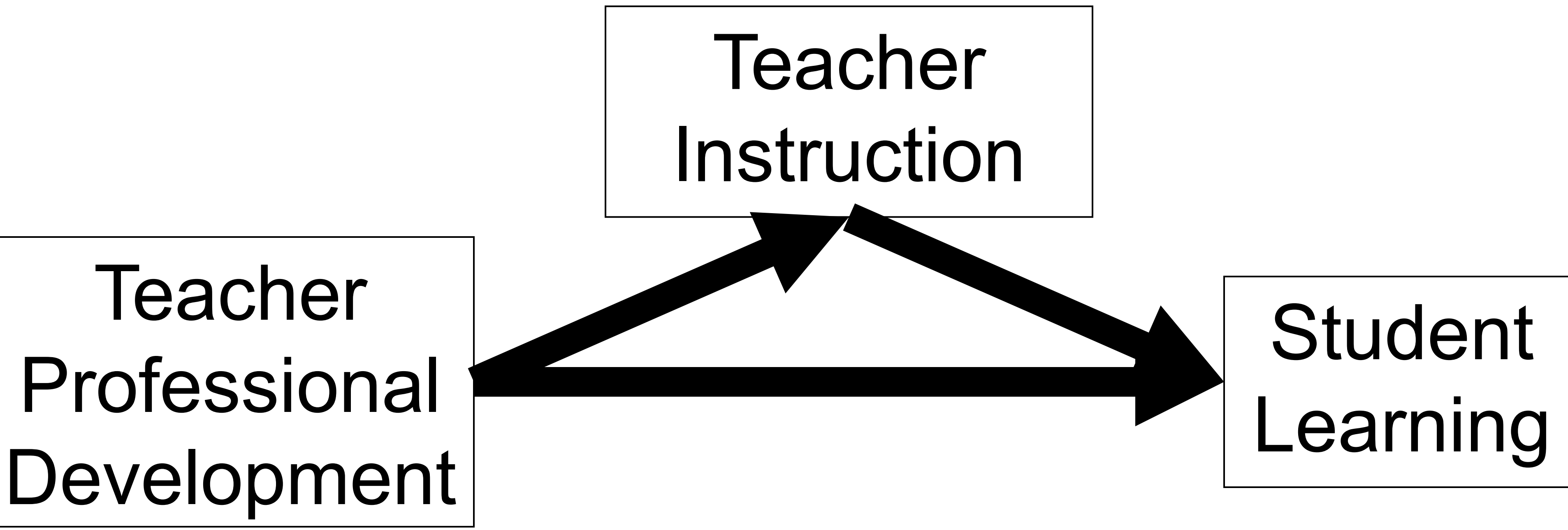
Without valid assessments of instructional process, it would be impossible to distinguish between these two explanations, severely limiting the study's contribution to knowledge. One would never know whether the theory underlying the program had in fact been tested.

Suppose instead that assignment to the innovation did produce gains in student learning. One might then assume that the innovation "worked" by improving instruction in the ways the program designers intended. But without valid measurements of instruction, this conclusion would be unwarranted. Perhaps the innovation "worked" in other ways, an assertion that could not be probed without studying the impact of the innovation on instruction. Once again, a failure to measure key aspects of classroom life yields major ambiguities in the findings.

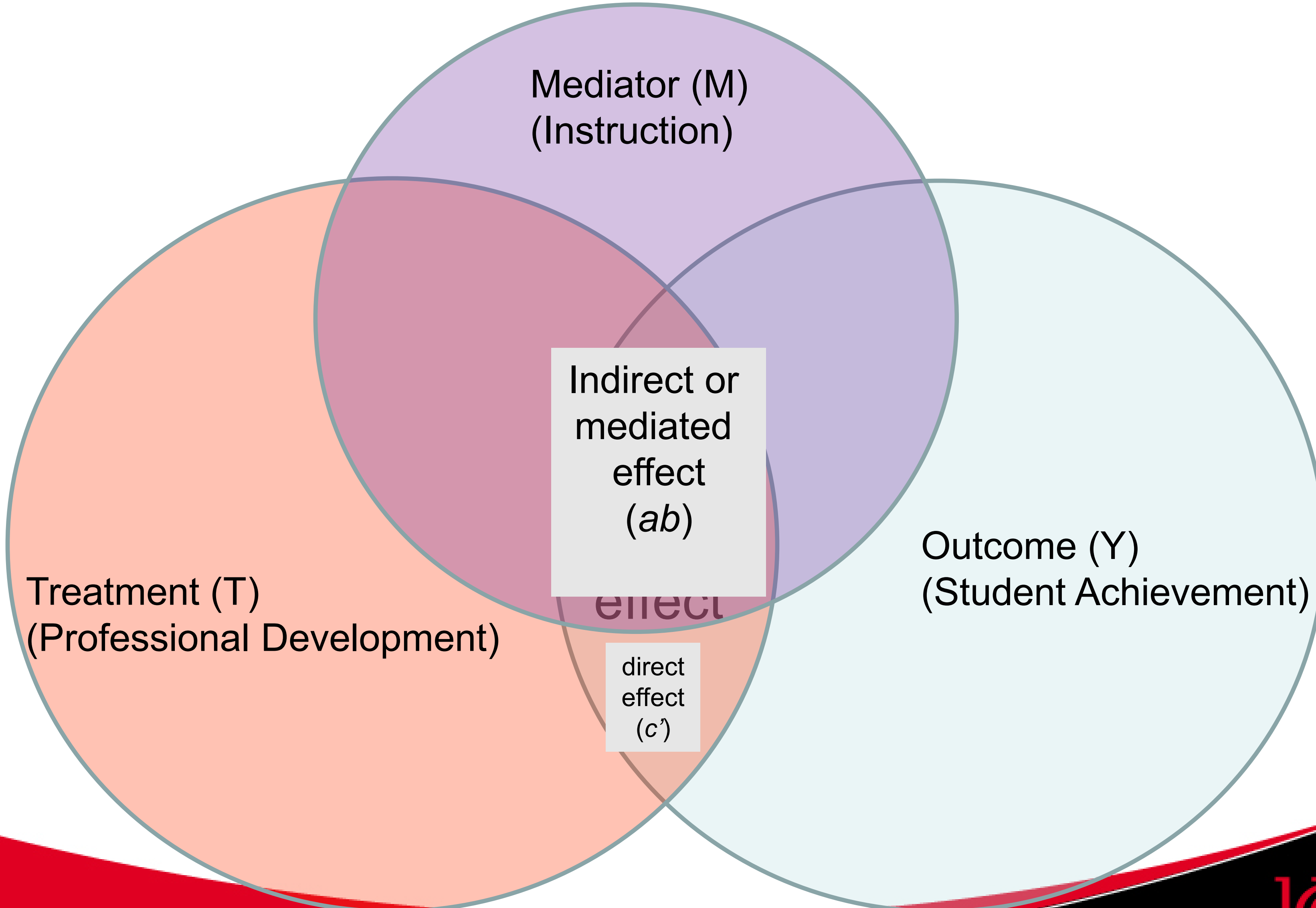
Mediation


Sample of Schools
Mediator
Measuring the process helps to open up the blackbox
Assess
Compare
Theory failure or implementation failure?

Example


Teacher Instruction
Teacher Professional Development
Student Learning

Mediation

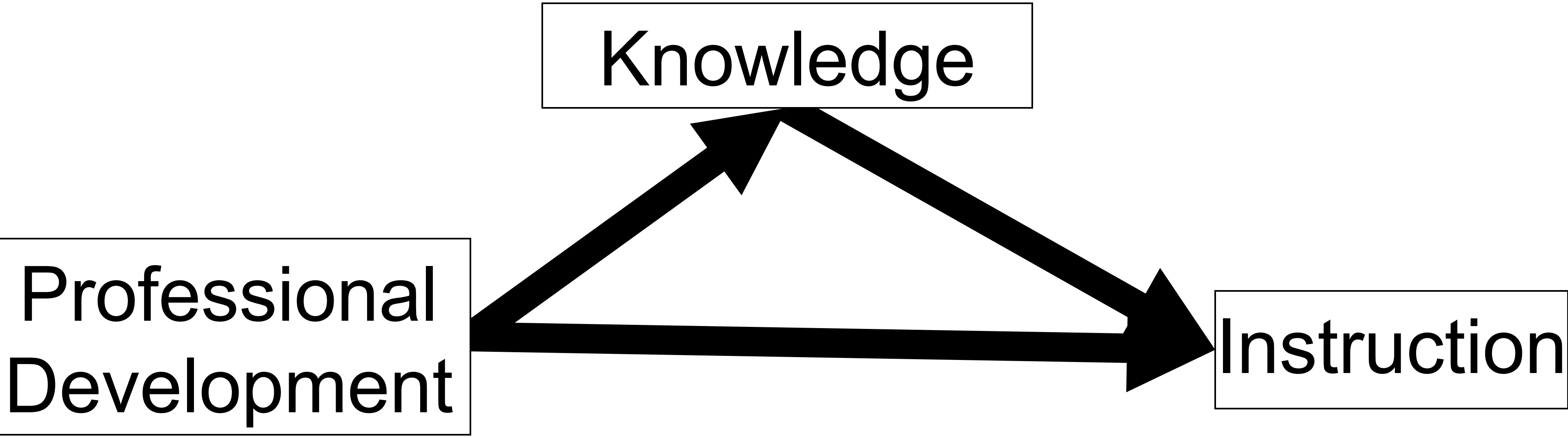


Mediator (M)
(Instruction)
Indirect or mediated effect (ab)
Treatment (T)
(Professional Development)
Outcome (Y)
(Student Achievement)
direct effect (c')

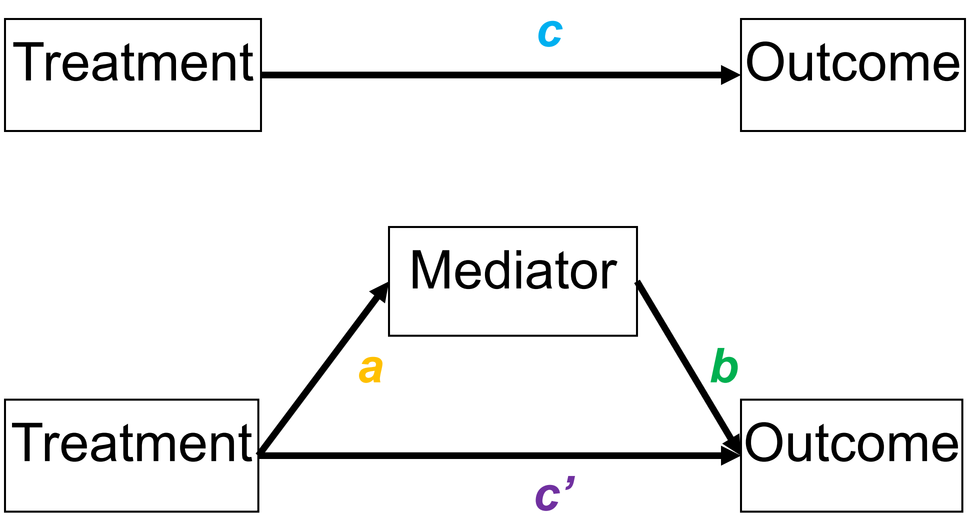
Examples

- Whole school intervention $\rightarrow$ *Coordination* $\rightarrow$ Achievement
- Therapy $\rightarrow$ *Engagement* $\rightarrow$ Outcome
  - E.g., Patient engagement
- Treatment $\rightarrow$ *Attitude* $\rightarrow$ Outcome
  - E.g., Motivational interviewing
- Professional Development $\rightarrow$ *Instruction* $\rightarrow$ Achievement
- Professional Development $\rightarrow$ *Knowledge* $\rightarrow$ *Instruction* $\rightarrow$ Achievement

Single Level Example



Simple single level mediation



$$Y_i = \beta_0 + c T_i + e_i$$

$$M_i = \beta_0 + a T_i + e_i$$

$$Y_i = \beta_0^{'} + b M_i + c^{'} T_i + e_i^{'}$$

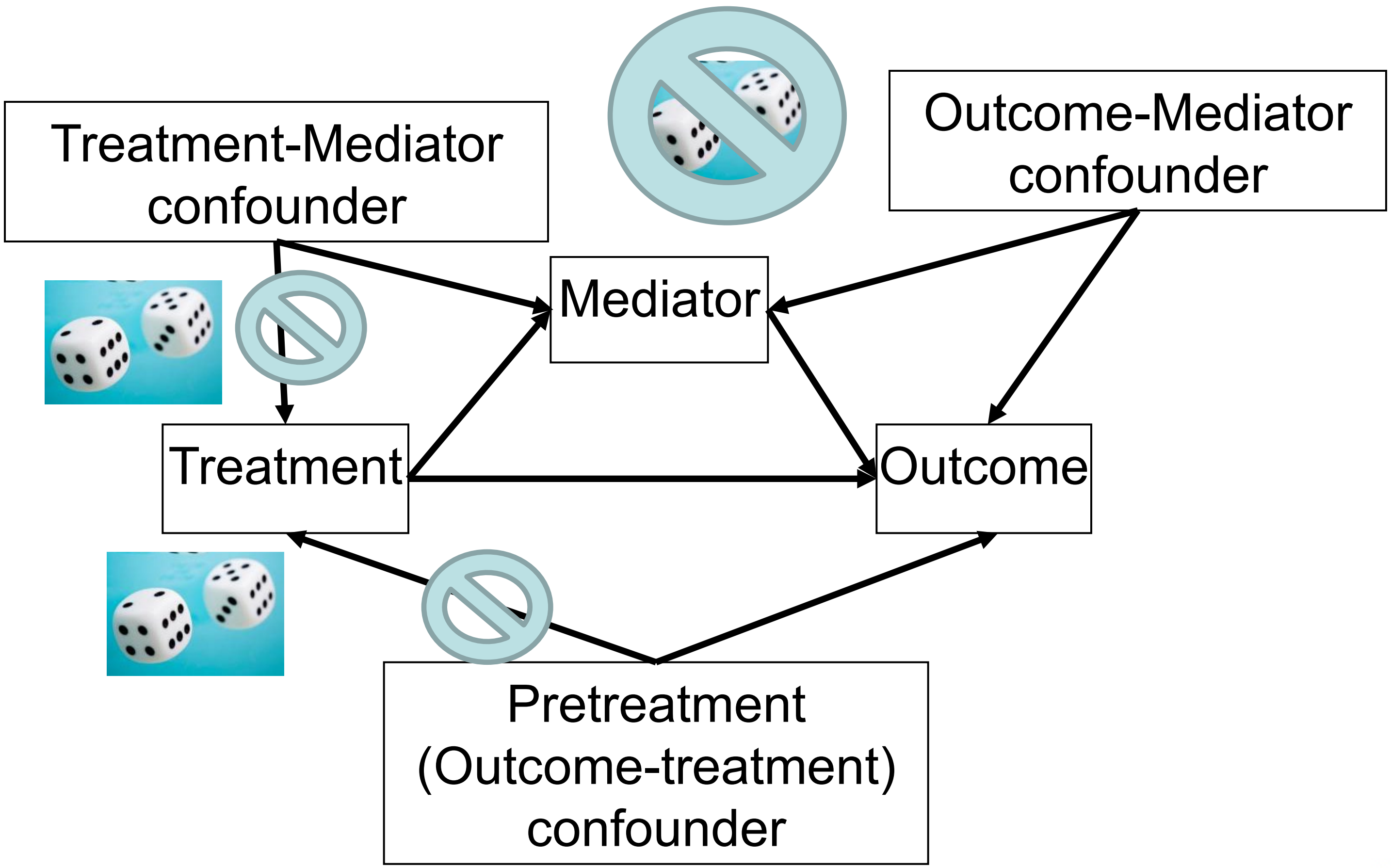
total effect = indirect effect + direct effect

c = ab + c’

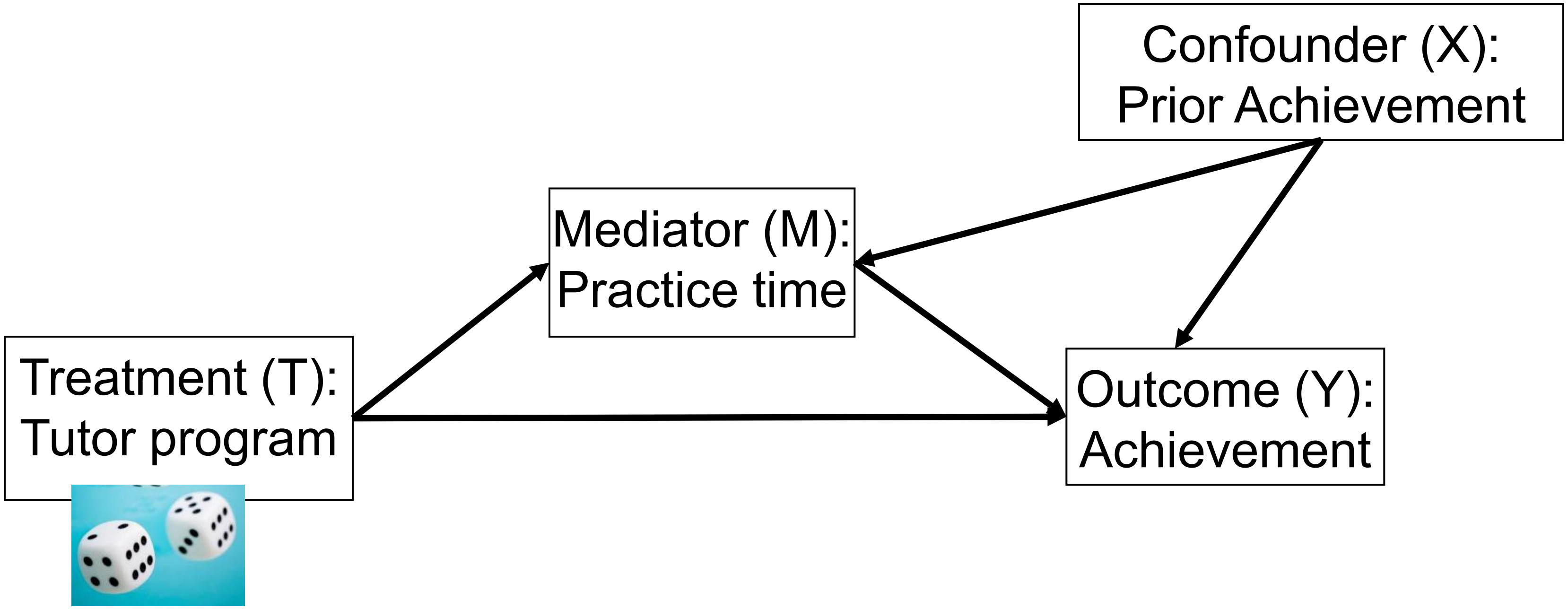
Assumptions: Sequential Ignorability

- Historical literature has generally treated estimates of mediation as causal
  - Causal inference regarding mediation requires that BOTH the treatment and mediator be randomly assigned or are ignorable
  - Random assignment of treatment ensures there are no pretreatment confounders that explain the observed outcome differences
  - Random assignment of the mediator ensures there no outcome-mediator confounders

Sequential Ignorability



Sequential Ignorability



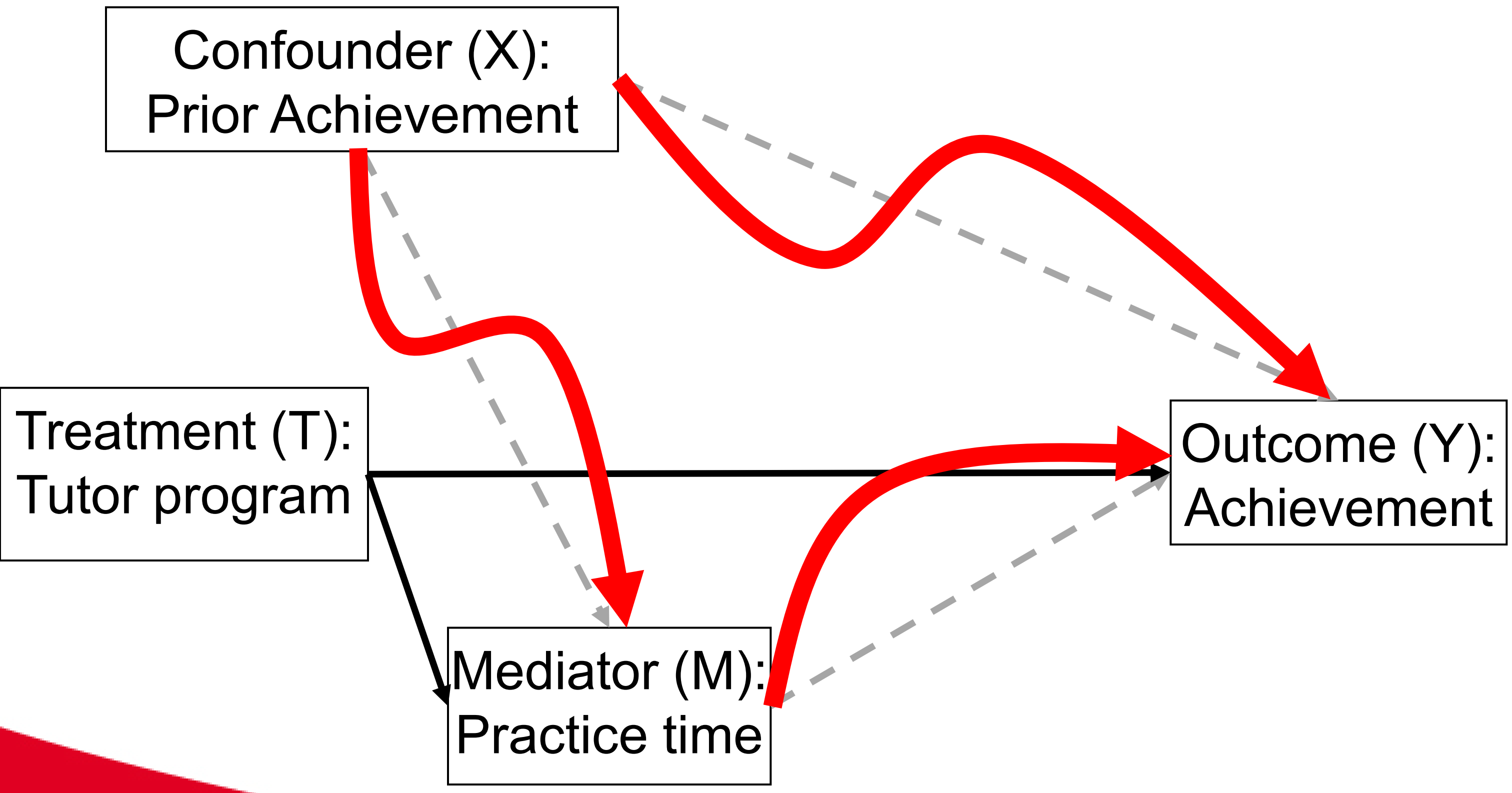
$$M_i = \beta_0 + aT_i \qquad + e_i$$

$$Y_i = \beta_{0Y} + bM_i + c'\ T_i \qquad + e_{iY}$$

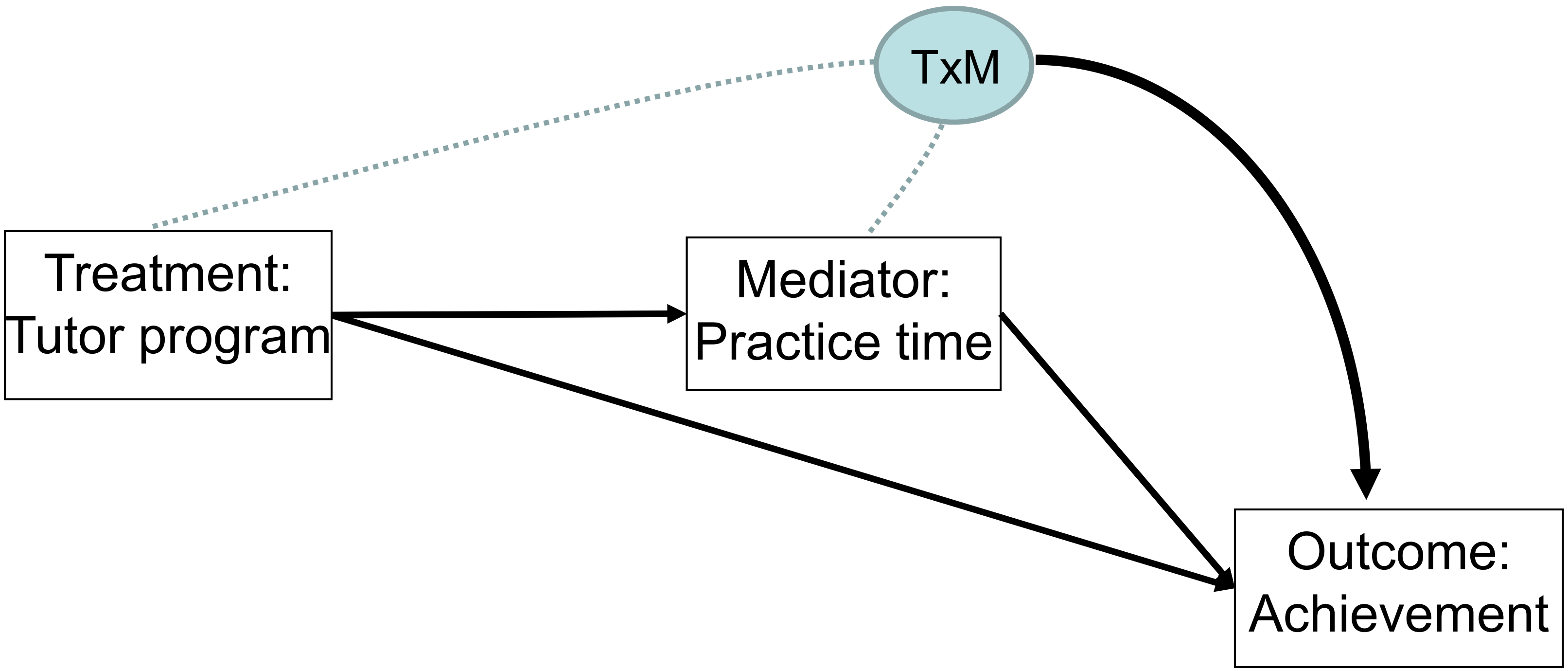
Assumptions: Correct Specification

- Ignorability also includes correct model specification
- Controlling for a background variable does not necessarily ensure ignorability
- Inferences are sensitive and susceptible to specification bias

Sequential Ignorability: Correct Specification



Correct Specification

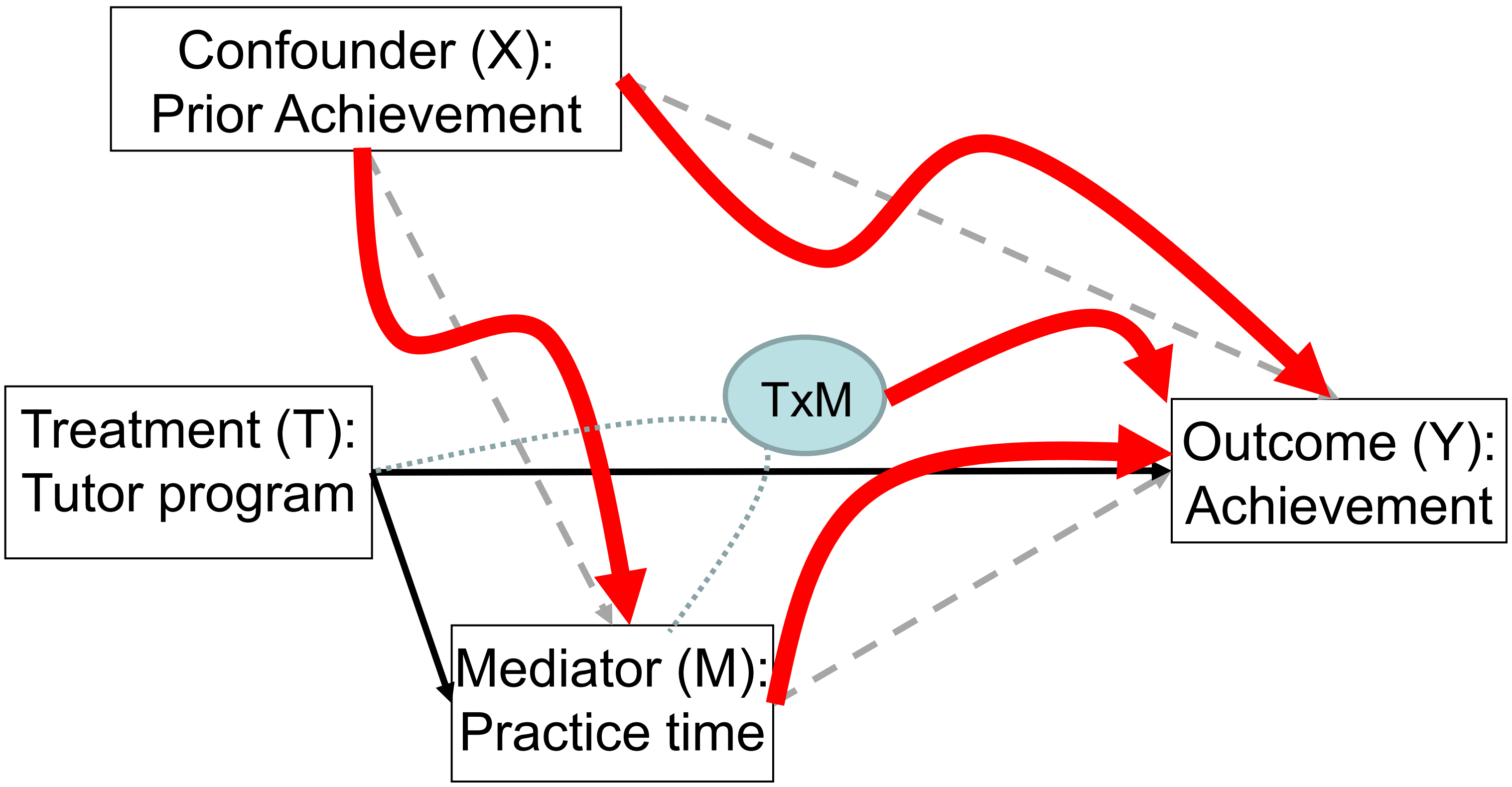


$$M_i = \beta_0 + aT_i + e_i$$

$$Y_i = \beta_{0Y} + bM_i + c'\ T_i + \beta_1 M_i T_i + e_{iY}$$

Implications

- Considerations:
  - Need ignorability for unbiased estimate
  - Models require precise specification
  - But typically only have coarse theories about relations, e.g.,

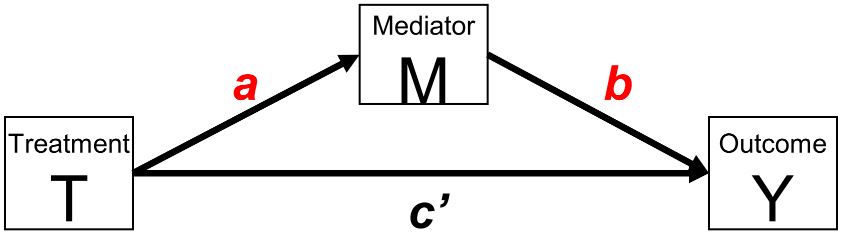


- **-Implication: Significant potential for model misspecification & incorrect estimates**

What to do? Machine Learning Approaches

- Model free definitions of effects (e.g., through the potential outcomes) provide productive ways to integrate machine learning and causal inference in ways that strengthen and relax assumptions
- Causal ML integrates prediction and explanation
  - Prediction and explanation are used as synergistic rather than competing tools
  - Prediction empirically establishes functional forms of relationships
  - Explanation compares counterfactuals

Testing for mediation effects



- ***ab*** is an estimate of the mediation effect (when there is not an interaction between T, M)
- Several possible approaches to test whether ab is nonzero and they differ in their power and type 1 error rates

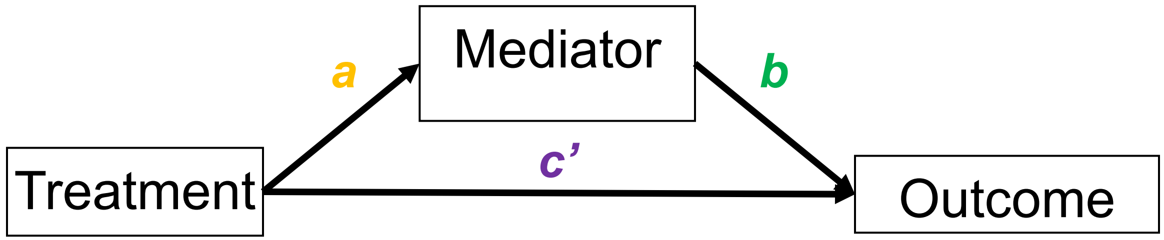
Tests of Mediation

- Some Common Tests of mediation
  - Test of joint significance
  - Monte Carlo confidence intervals (a form of parametric bootstrapping)
  - Bootstrap resampling (e.g., non-, semi-parametric)
  - Sobel test

Joint significance approach

1) Test a path
2) Test b path

*If both are significant then infer mediation



--Power is good approximation to more complex methods

--Type I error rate slightly lower than expected

Monte Carlo Confidence Interval Test

- Estimates empirical distribution of the ab product using resampling based methods (type of parametric bootstrapping)
  - Draw samples of *a* and *b* paths from their respective distributions, multiply them and repeat to approximate the posterior distribution of their product *ab*
  - If CI does not include zero, then infer mediation
- Does not require full data (useful for design purposes)
- Does not assume any sampling distribution of the indirect effect or that it is symmetric
- Typically found to be very powerful and comparable to other bootstrap based intervals

Mediation in
Cluster Randomized Trials

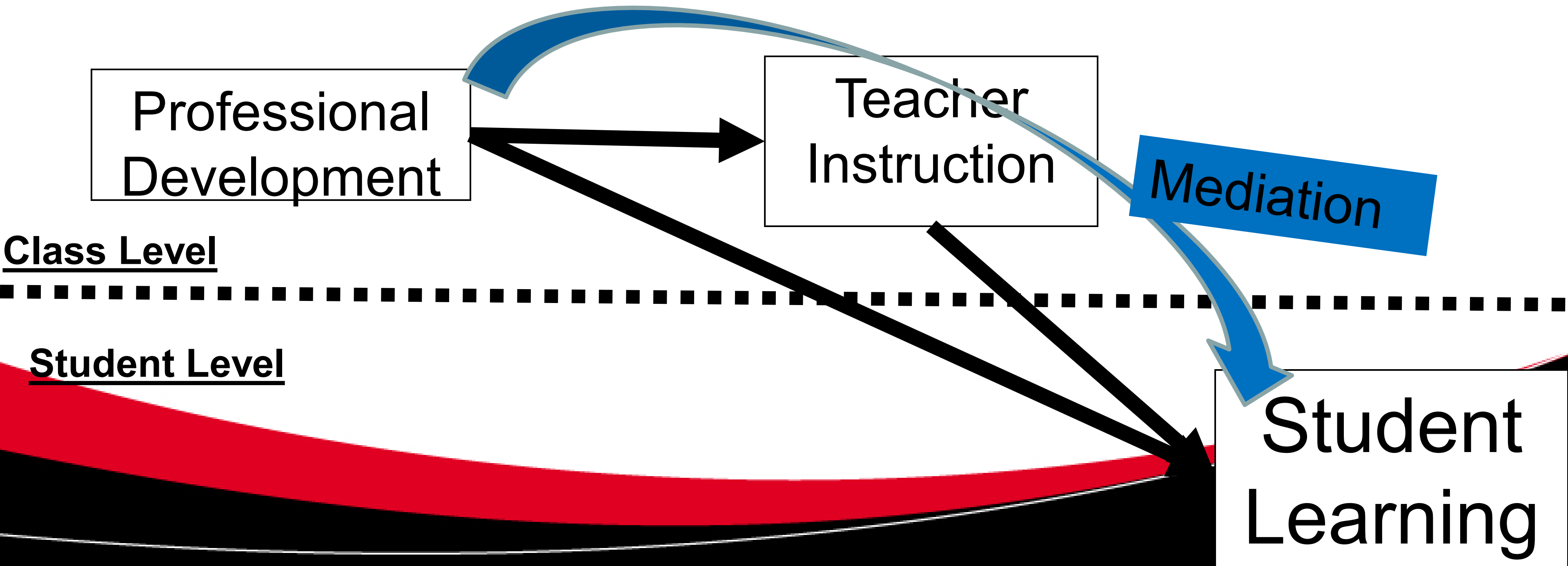
UNIVERSITY OF Cincinnati

Mediation in Cluster Randomized Trials

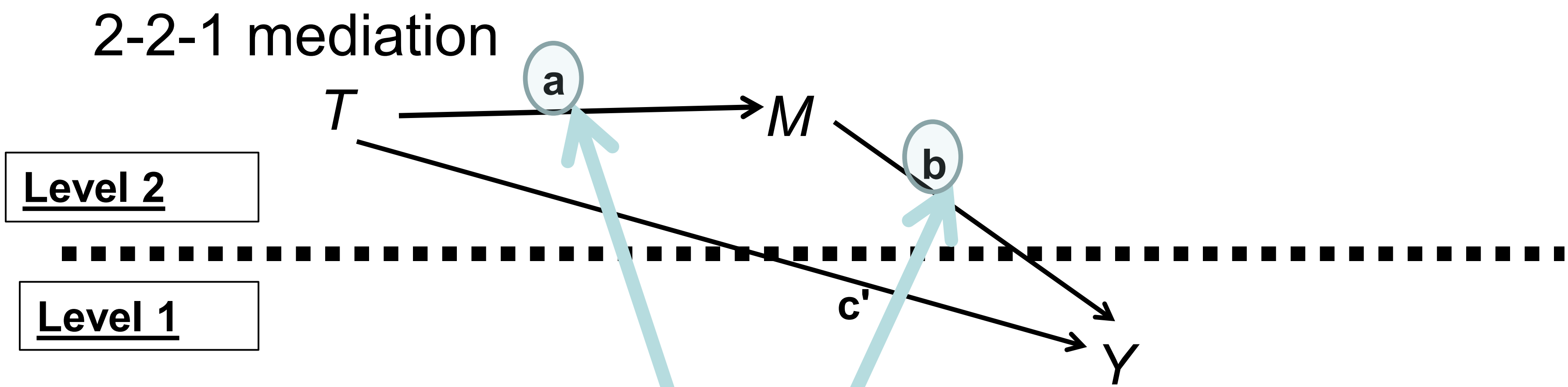
- Lots of combinations and models depending on level of each focal variable: $T \rightarrow M \rightarrow Y$,
- *e.g.*,
  - Upper-level mediation [$2 \rightarrow 2 \rightarrow 1$]
  - Cross-level mediation [$2 \rightarrow 1 \rightarrow 1$]
  - Three level and sequential mediation versions e.g., [$3 \rightarrow 2 \rightarrow 2 \rightarrow 1$] when considering how school randomized professional development programs impact student achievement via instruction (via teacher knowledge)

Simple 2-2-1 Mediation

- Imagine a cluster randomized trial that assigns teachers to receive professional development (PD) or a business as usual control condition. The aim of this PD is to improve the teachers' quality of instruction (IQ) so that, in turn, students' achievement (Y) increases.



2-2-1 Mediation Model



Mediation model

Class level $M_j = \pi_0 + aT_j + \varepsilon_j^M \qquad \varepsilon_j^M \sim N(0, \sigma_{M|}^2)$

Outcome model

Student level $Y_{ij} = \beta_{0j} + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N(0, \sigma_{Y|}^2)$

Class level $\beta_{0j} = \gamma_{00} + bM_j + c'T_j + u_{0j} \qquad u_{0j} \sim N(0, \tau_{Y|}^2)$

Example in R

- 2-2-1 Data at

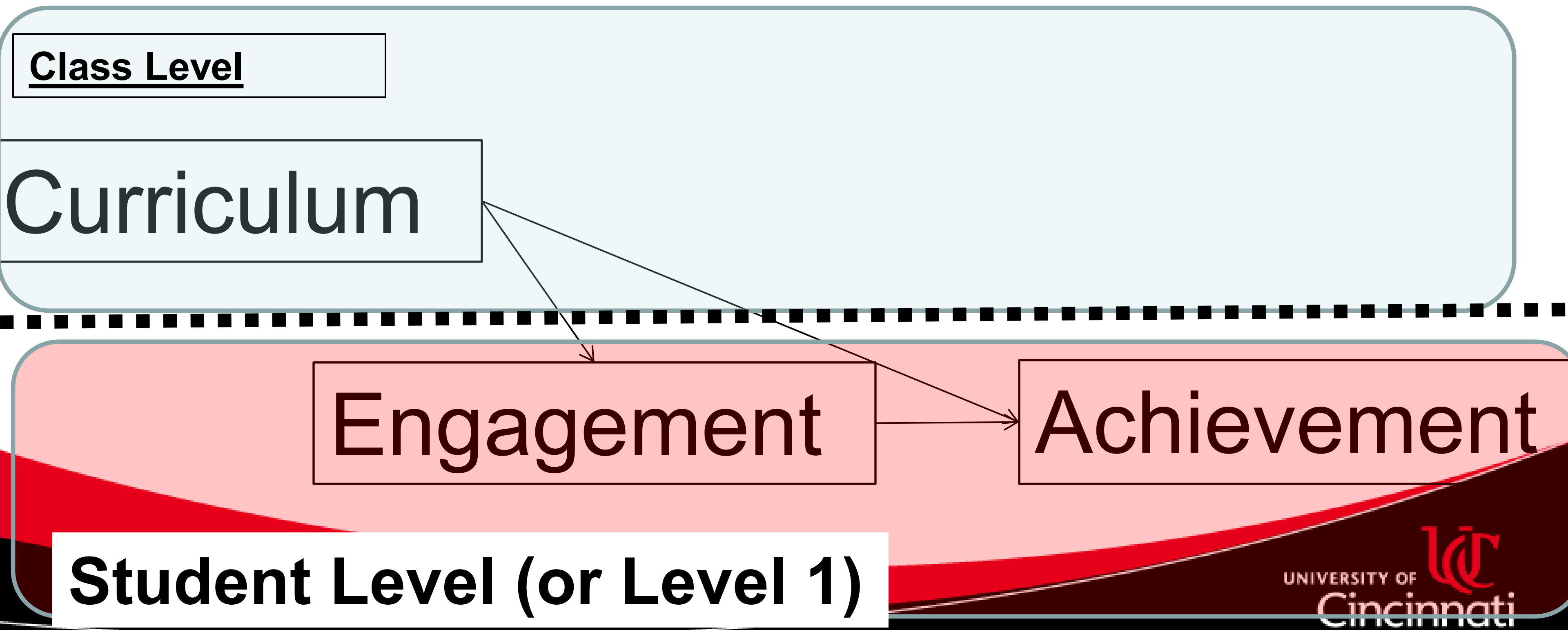
tinyurl.com/ypnx24z6

- Syntax file at

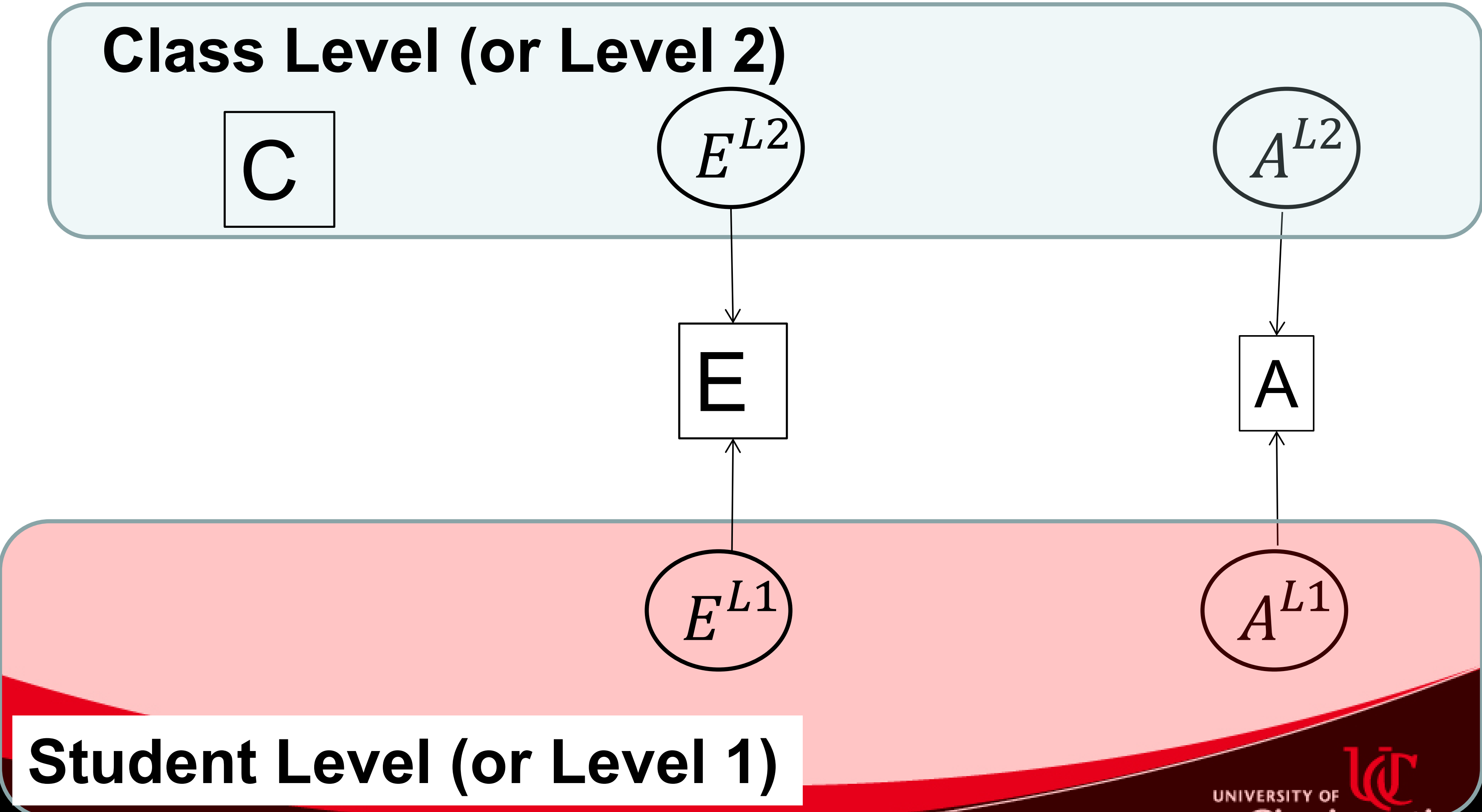
https://tinyurl.com/23fexush

Simple 2-1-1 Mediation

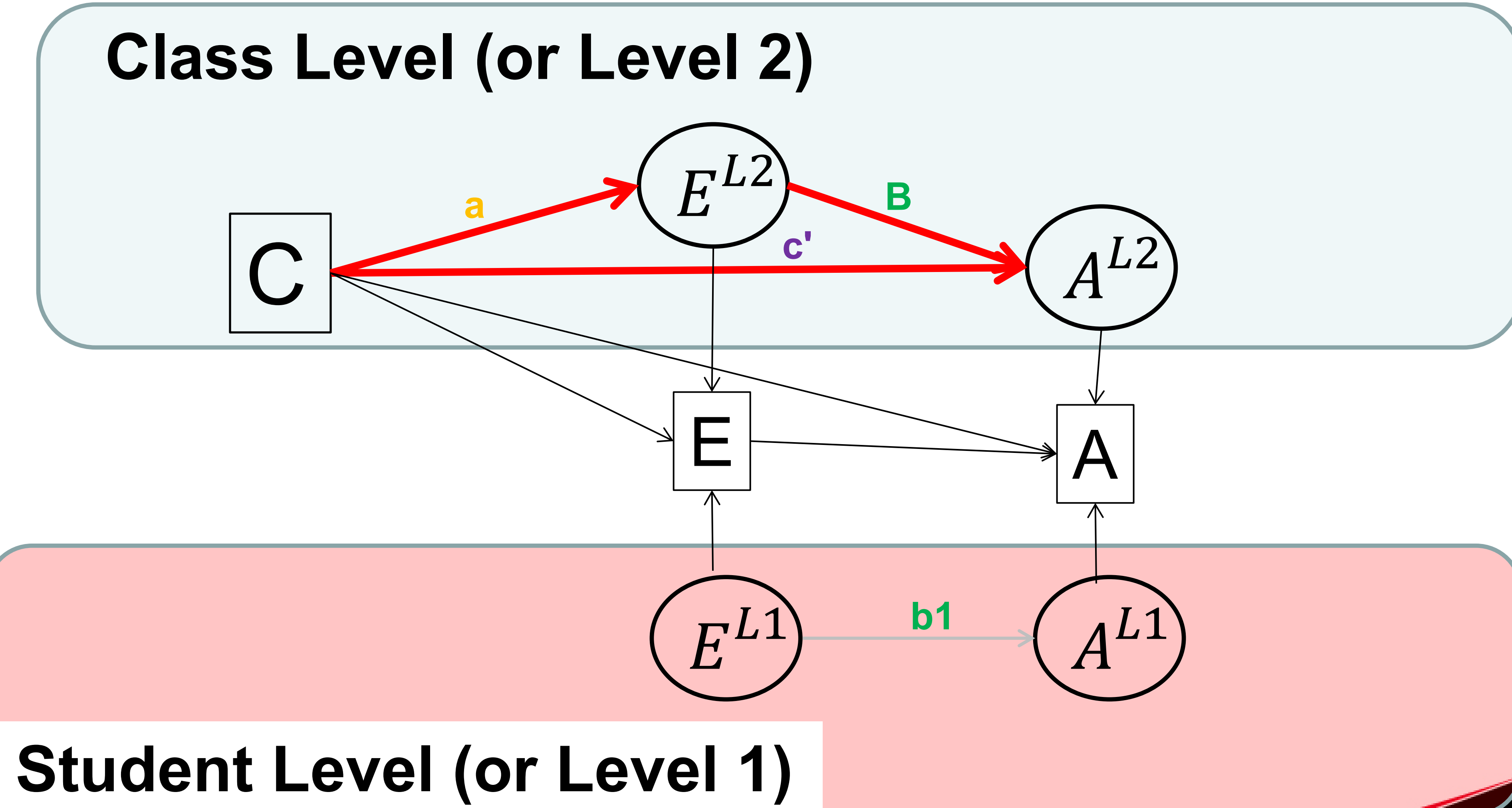
Imagine a cluster randomized trial that assigns teachers to use a new curriculum or the conventional curriculum. The theory of action underlying the curriculum is that it will improve student engagement which will in turn improve achievement.



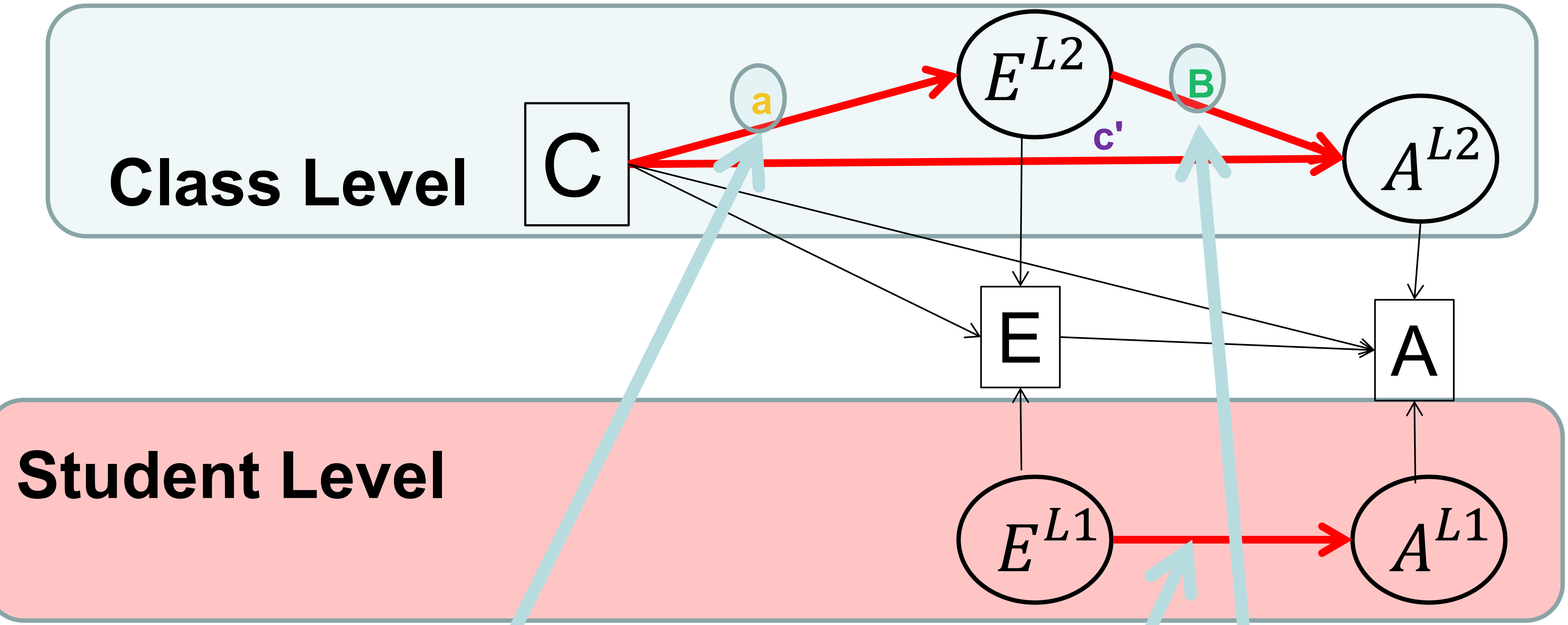
Multilevel Decomposition of Endogenous Variables



Classroom Engagement



Classroom Engagement



Mediator Model

$$E_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + a C_j + u_{0j}$$

Outcome Model

$$Y_{ij} = \beta_{0j} + b_1 E^{L1} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + c' \; C_j + B \, E^{L2} + u_{0j}$$

Example in R

- 2-1-1 Data at

https://tinyurl.com/55wmyhpw

Break

- Questions, Comments, & Feedback
  ben.kelcey@gmail.com

From Analysis to Design: Planning Studies to Examine Mediation

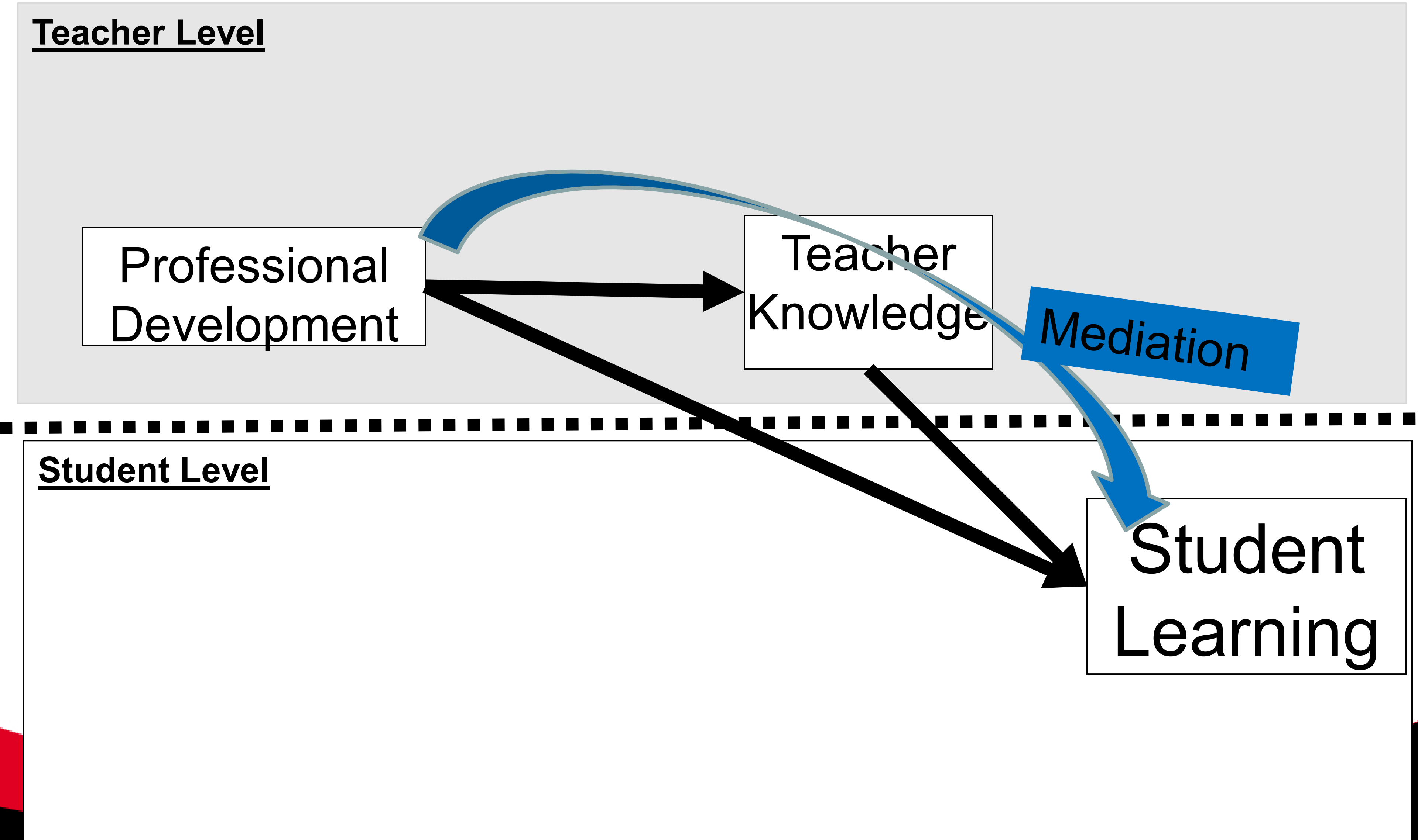
From Analysis to Design

- How might we design studies to ensure they have reasonable chance of detecting mediation effects if they exist?
  - E.g., what are reasonable sample sizes?
- What is the requisite scale for sufficiently powered studies targeting multilevel mediation?
- Are ‘typical’ sample sizes enough?

Power Analyses for Multilevel Mediation

- Simple two-level mediation example
  - Teachers are randomly assigned to participate in a PD program designed to equip teachers with core pedagogical and substantive knowledge
  - Students nested within teachers
- Outcome of interest is student achievement
- Mediator of interest is teacher knowledge
- Goal: Design a study to detect if the impact of PD on student achievement is mediated by changes in teacher knowledge

Multilevel Mediation (2-2-1)



Parameters Governing Power for 2-2-1

Same as for Main Effect

J : total number of clusters

n : number of individuals per cluster

ρ : Intraclass correlation coefficient

R_{L1}^2 : proportion of variance explained at level-1

R_{L2}^2 : proportion of variance explained at level-2

P: proportion of level-2 units randomized to treatment

New Parameters for Mediation

a: treatment-mediator path coefficient

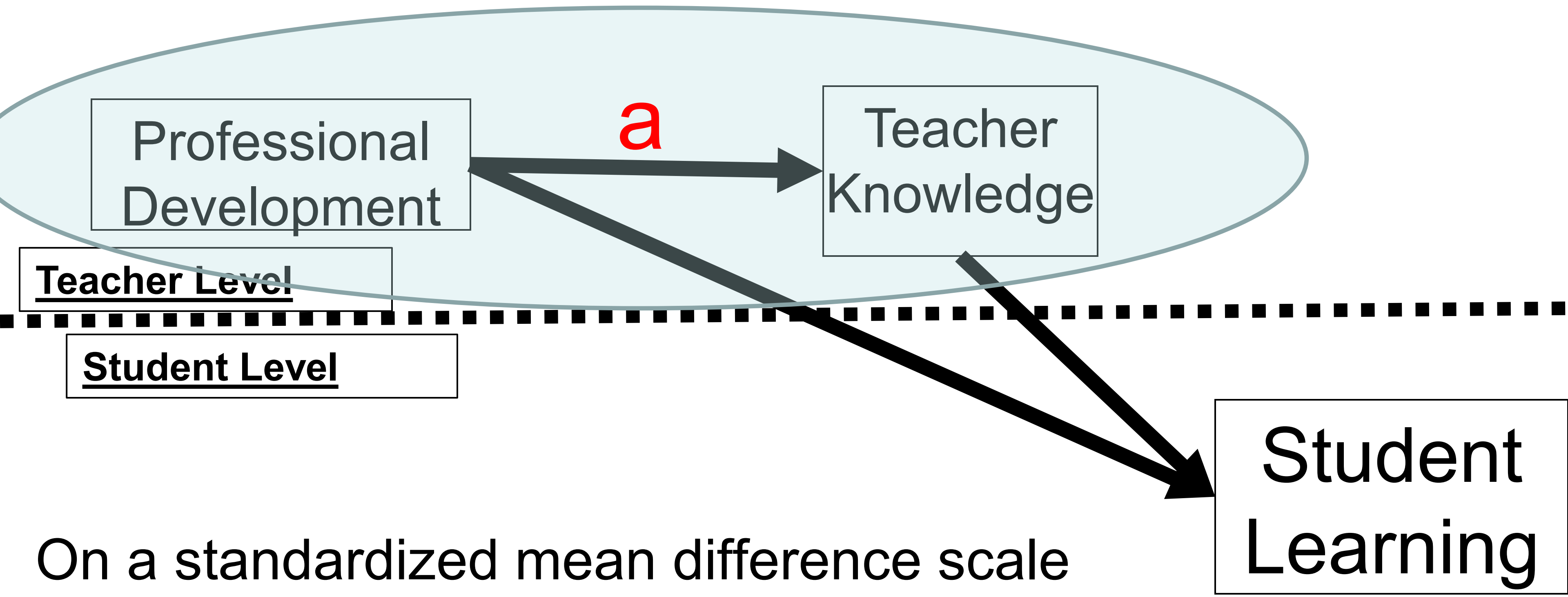
b: mediator-outcome path coefficient

c': direct effect of treatment on outcome

R_M^2 : proportion of mediator variance explained by covariates

Cluster-Level Mediation a Parameter

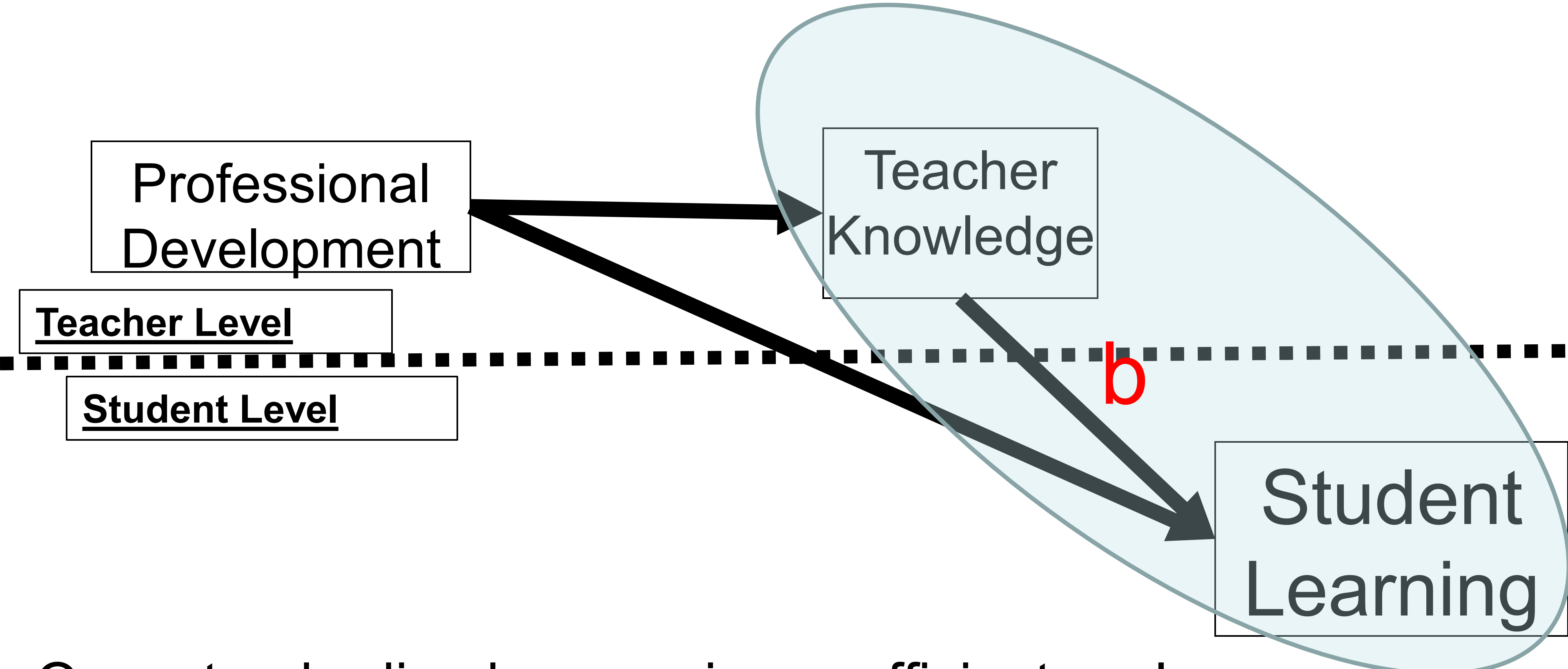
a: treatment-mediator path coefficient



On a standardized mean difference scale (same as main effect)

Cluster-Level Mediation
b Parameter

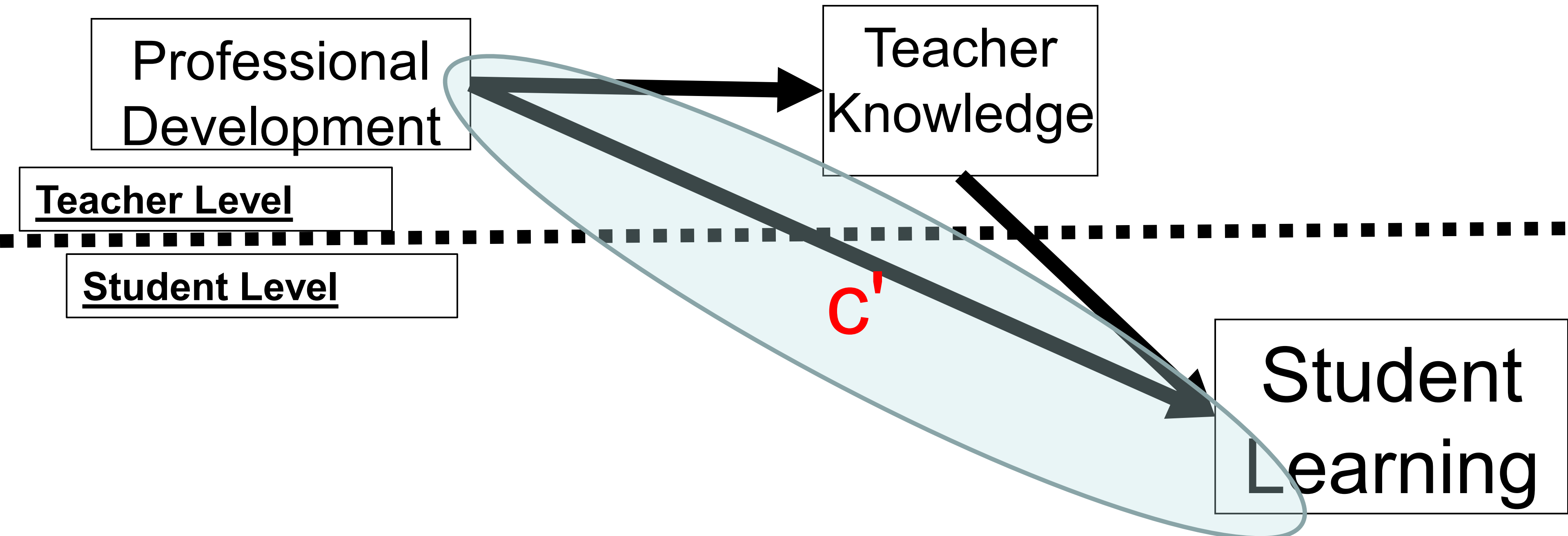
b: mediator-outcome path coefficient



On a standardized regression coefficient scale

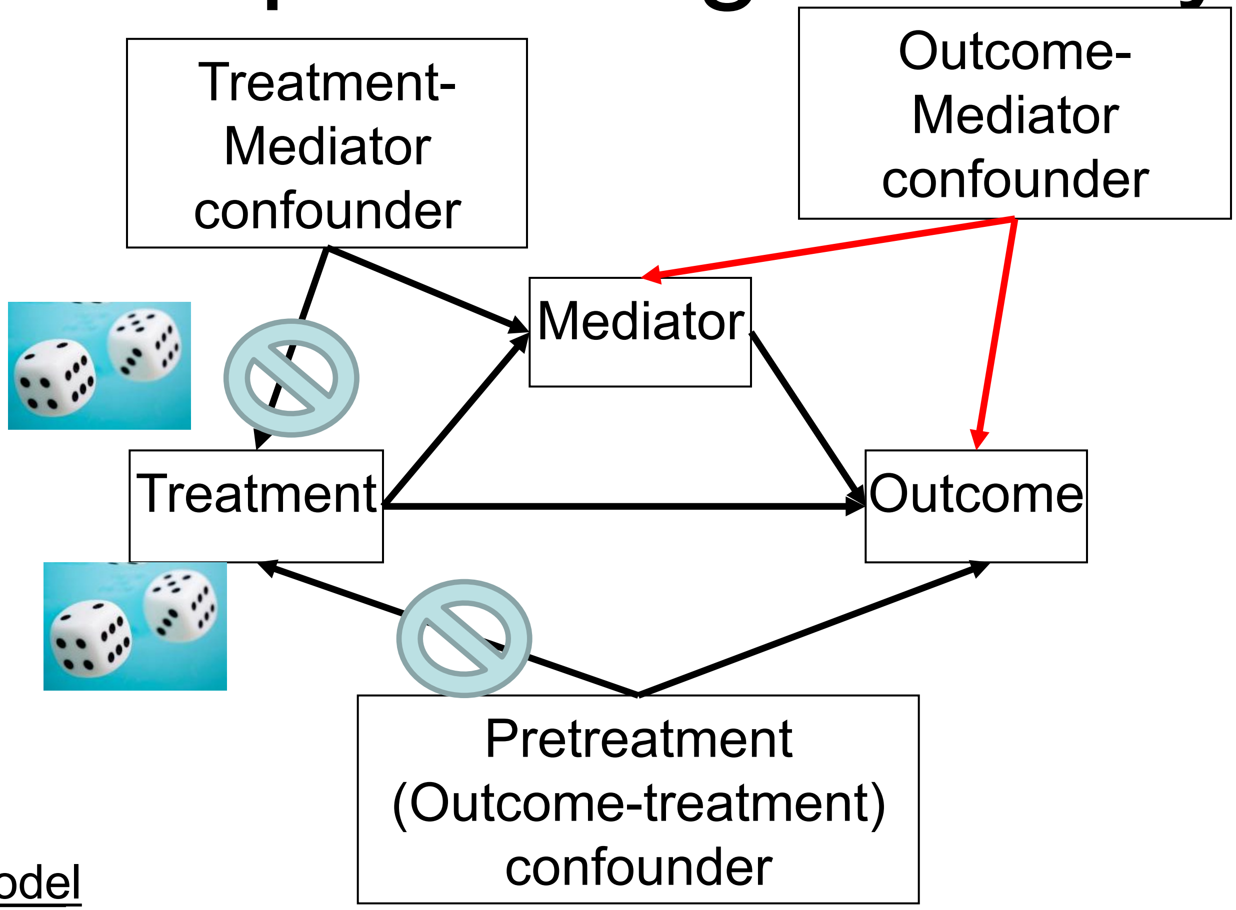
Cluster-Level Mediation
c’ Parameter

c‘: direct effect of treatment on outcome



On a standardized mean differences scale

Sequential Ignorability



<u>Mediation model</u>

Class level

$$M_j = \pi_0 + aT_j + \beta_1 X_j^{L2} + \varepsilon_j^M \qquad \varepsilon_j^M \sim N(0, \sigma_{M|}^2)$$

<u>Outcome model</u>

Student level

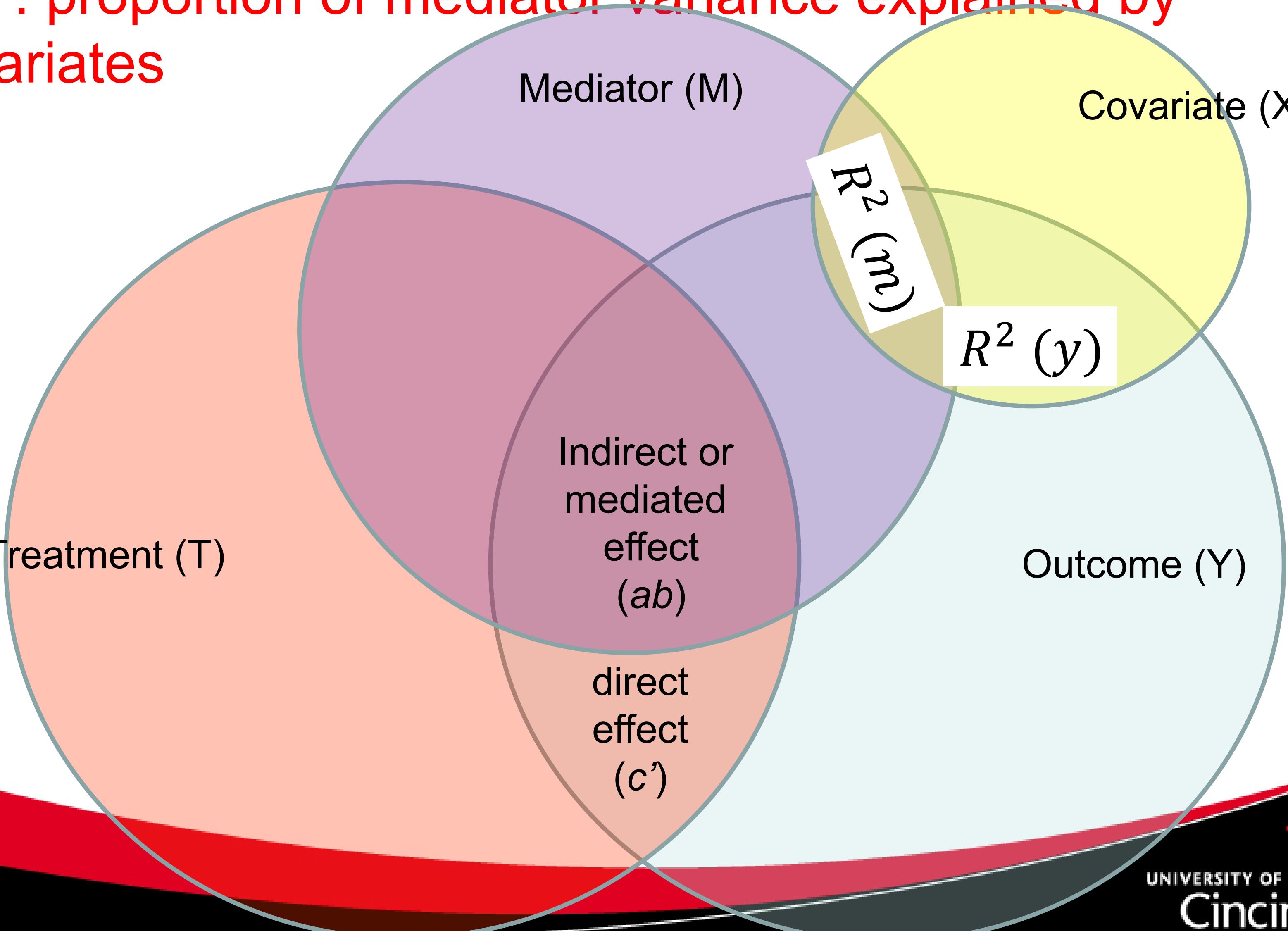
$$Y_{ij} = \beta_{0j} + \beta_{1Y} X_{ij}^{L1} + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N(0, \sigma_{Y|}^2)$$

Class level

$$\beta_{0j} = \gamma_{00} + bM_j + c'T_j + \gamma_1 X_j^{L2} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{Y|}^2)$$

Cluster-Level Mediation R_M^2 Parameter

R_M^2 : proportion of mediator variance explained by covariates



Scale of Effect Size

- Lots of different approaches
  - Review lit and identify most meaningful for your context

- One simple approach: Multiply *a* and *b* paths where the magnitude of the paths is based on common (theoretical or empirical) effect size interpretations

Effect Size

- *a* path
  - standardized mean difference scale for dichotomous treatments
- *b* path
  - If the mediator and outcome are standardized, its on a standardized regression coefficient scale (controlling for treatment and covariates)
- Then effect size is just product of a and b

Possible (theoretical) Benchmarks

- Dichotomous treatment, continuous mediator

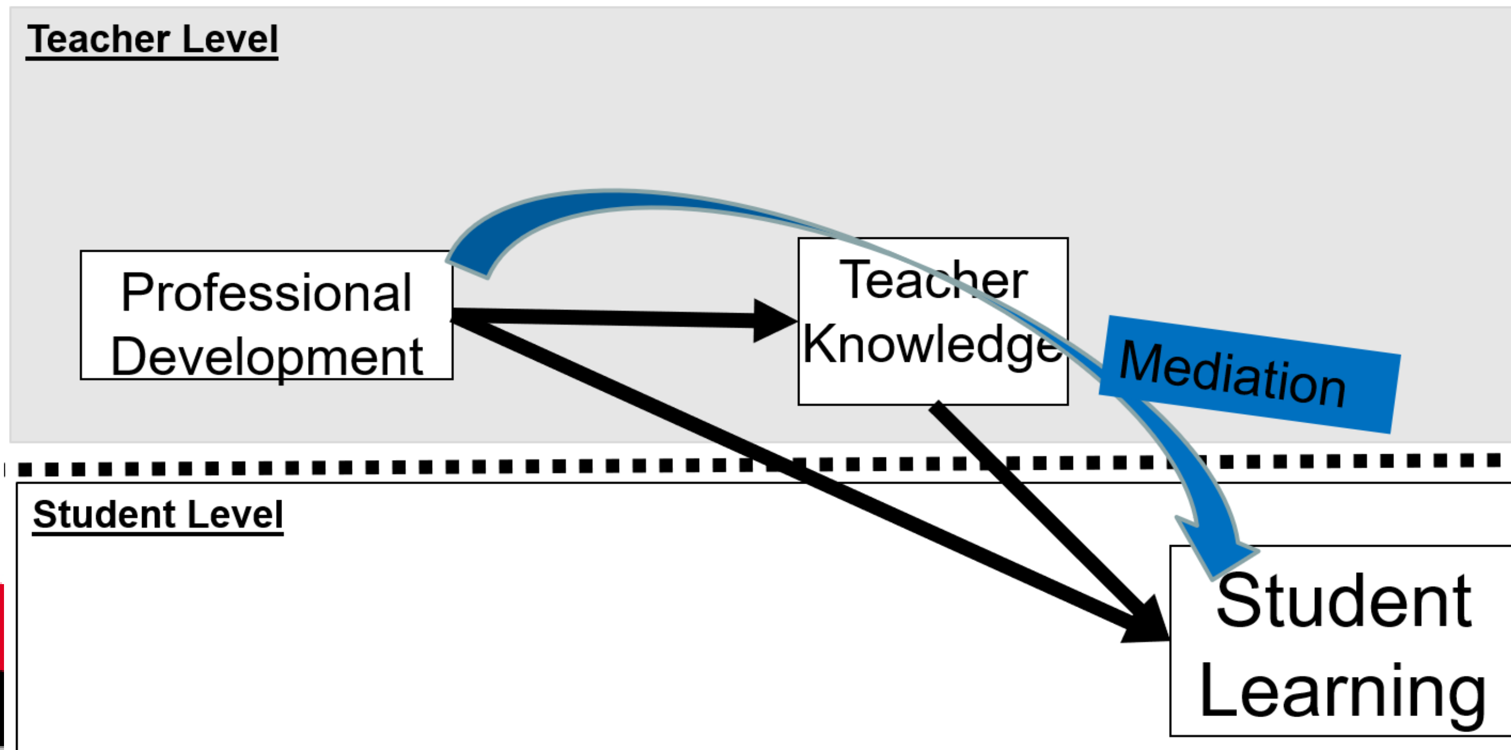
Size	= XX (i.e., effect of a * effect of b)
– Small	= .02 (i.e., .2*.1)
– Medium	= .15 (i.e., .5*.3)
– Large	= .40 (i.e., .8*.5)

Tests of Mediation

- Some Common Tests of mediation
  - Sobel test
  - Test of joint significance
  - Monte Carlo interval test
  - Bootstrap resampling

Example Power Analysis

Consider a professional development program that aims to improve student learning by improving teacher knowledge. Assume teachers are randomly assigned to participate in the professional development program or a control condition. If we plan to sample about 20 students per teachers, how many teachers do we need for an 80% chance of detecting a mediation effect? (more info on next slide)



Example: Cluster-Level Mediation Parameters

a: treatment-mediator path coefficient — 0.5

b: mediator-outcome path coefficient — 0.3

c‘: direct effect of treatment on outcome — 0.1

ρ : Intraclass correlation — 0.15

$R_{L_1}^2$ (R21): proportion of variance explained at level-1 — 0.5

$R_{L_2}^2$ (R22): proportion of variance explained at level-2 — 0.5

R_M^2 (R2m2): proportion of mediator variance explained by covariates — 0.5

P: proportion of level-2 units randomized to treatment — 0.50

n : number of individuals per cluster — 20

J : total number of clusters — ??

PowerUpR Shiny App

Web-based Shiny app
https://powerupr.shinyapps.io/index/
Manual for Shiny app
https://www.causalevaluation.org/power-analysis.html

Excel version
https://www.causalevaluation.org/power-analysis.html

R package
https://cran.r-project.org/web/packages/PowerUpR/index.html

Results


```
Statistical power:
------------------------------------
       t sobel joint  mc
a  0.793    NA    NA  NA
b  1.000    NA    NA  NA
ab    NA 0.751 0.793 0.8
------------------------------------
Degrees of freedom for path a: 52
Degrees of freedom for path b: 52
Standardized standard error for path a: 0.177
Standardized standard error for path b: 0.041
Type I error rate: 0.05
Two-tailed test: TRUE
```

UNIVERSITY OF Cincinnati

Exercise

Program: Consider a new school wide behavioral intervention program focusing on Positive Behavior Supports (PBS). The new program is to be implemented at the school wide and expected to impact student outcomes by improving school climate. The researchers plan to randomly assign schools to the treatment or control condition.

RQ1: To what extent does the new program improve behavioral outcomes (main effect)?

RQ2: To what extent does the program operate through changes in the school climate (mediation effect)?

Design: How large of a sample do we need to detect a mediation effect with 80% power?

Example parameter values on next slide...

2-2-1 Mediation Parameters

a: treatment-mediator path coefficient—0.5

b: mediator-outcome path coefficient—0.25

c‘: direct effect of treatment on outcome—0.1

ρ : Intraclass correlation—0.25

$R_{L_1}^2$: proportion of variance explained at level-1—0.4

$R_{L_2}^2$: proportion of variance explained at level-2—0.5

R_M^2 : proportion of mediator variance explained by covariates—0.3

P: proportion of level-2 units randomized to treatment—0.5

n : number of individuals per cluster—50

J : total number of clusters—??

Function

Effect

Mediator

Output

Statistical power

Design

>> Two-level CRT - Level 2 Mediator

Parameters

Standardized effect for path a (`esa`)

0.5

Standardized effect for path b (`esb`)

0.25

Standardized effect for path c prime (`escp`)

0.1

Type I error rate (`alpha`)

0.05

Two tailed test? (`two.tailed`)

TRUE

Proportion of variance in the outcome between level 2 units (ICC2) (`rho2`)

0.25

Proportion of variance in the outcome explained by level 1 covariates (`r21`)

0.4

Proportion of variance in the outcome explained by level 2 covariates (`r22`)

0.5

Proportion of variance in the mediator explained by level 2 covariates (`r2m2`)

0.3

(Average) proportion of units randomly assigned to treatment condition (`p`)

0.5

Sample Size

Level 1 sample size (`n`)

50

Level 2 sample size (`J`)

80

Controls

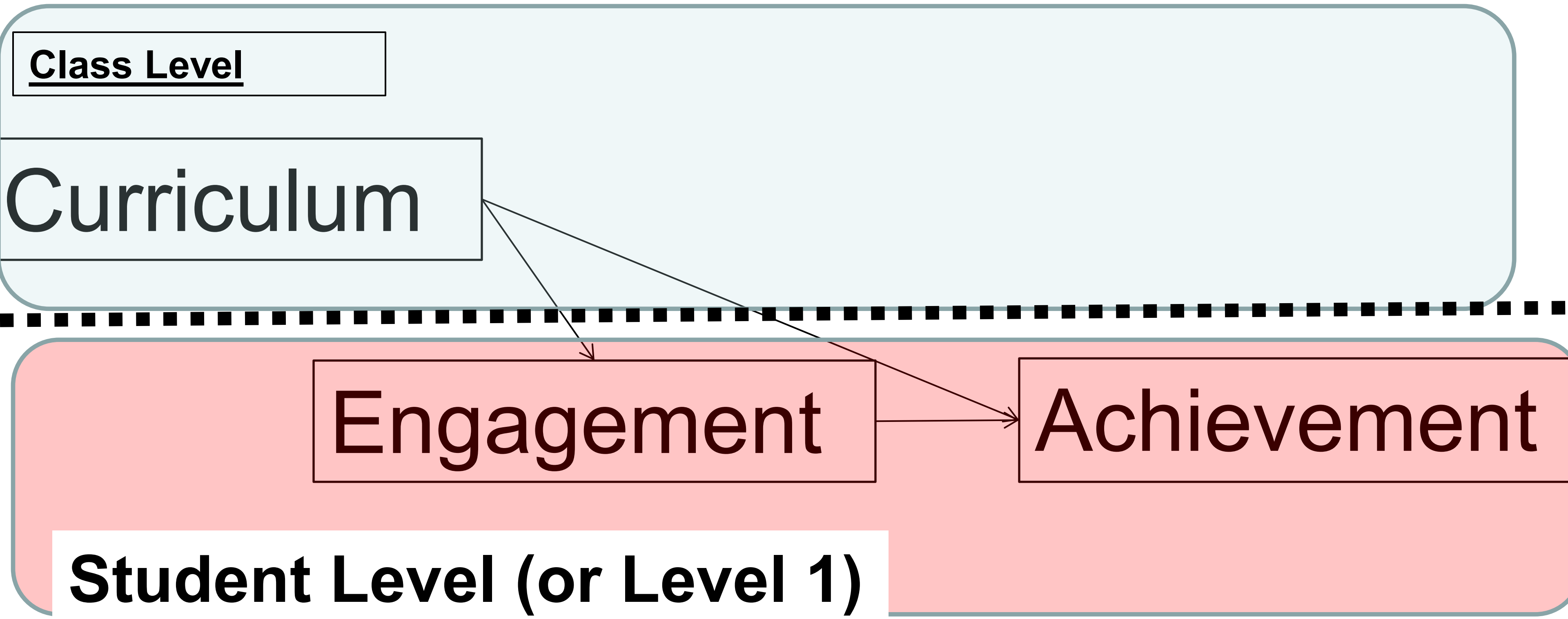
Update | Bookmark | Reset | Quit

```
Statistical power:
------------------------------------
        t sobel joint    mc
a  0.79    NA    NA    NA
b  1.00    NA    NA    NA
ab   NA 0.724  0.79 0.809
------------------------------------
Degrees of freedom for path a: 76
Degrees of freedom for path b: 76
Standardized standard error for path a: 0.179
Standardized standard error for path b: 0.04
Type I error rate: 0.05
Two-tailed test: TRUE
```

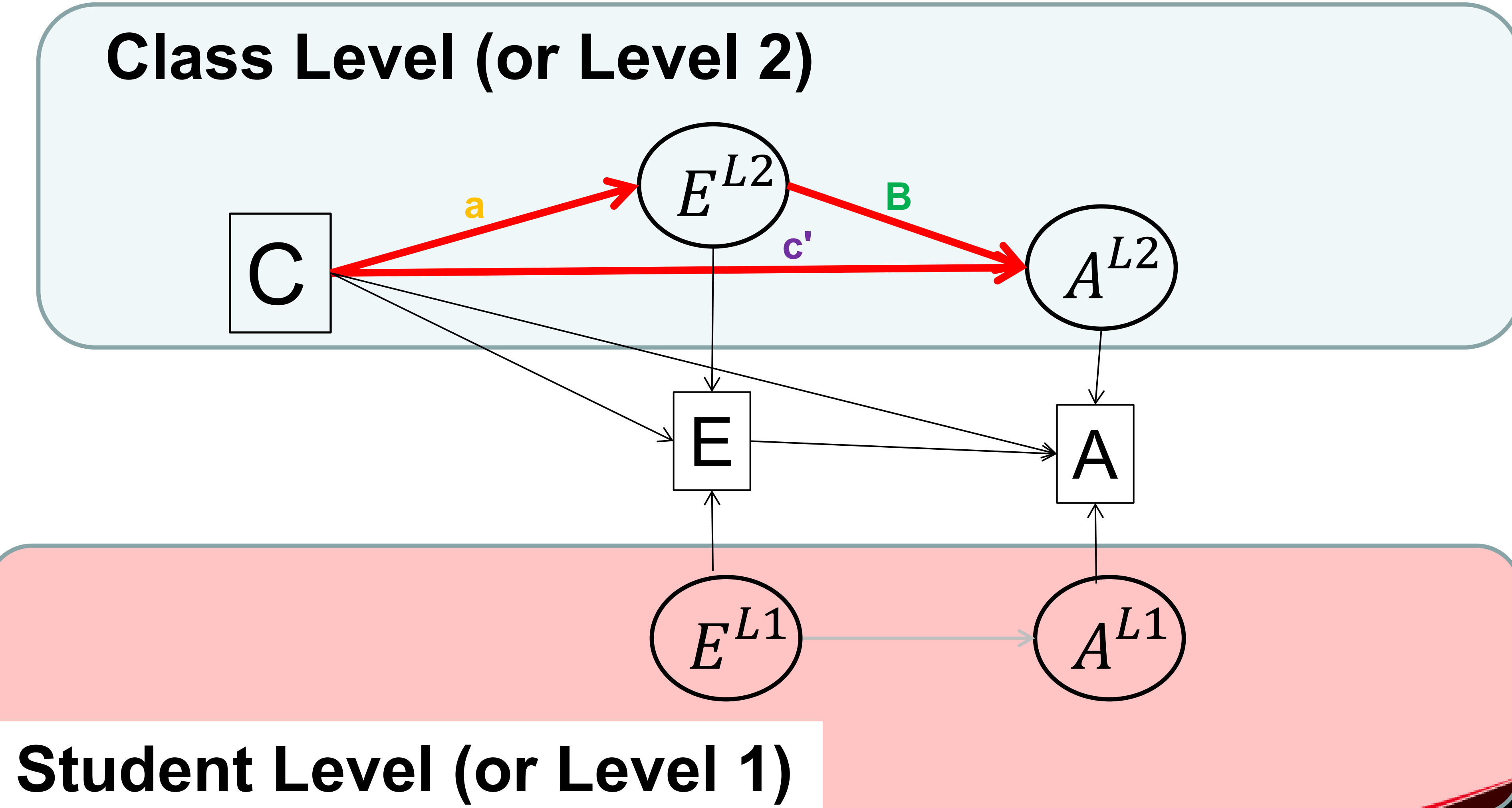
2-1-1 Power Analysis Example

- Consider a simple two-level mediation example with students nested within classes that are randomly assigned to participate in an innovative curriculum designed to engage students of all levels
- Let the outcome of interest be students’ achievement and assume that the mediator of interest is student engagement
- We are interested in designing a study to detect the extent to which the impact of participating in the innovative curriculum on student achievement is mediated by changes in (individual and collective) student engagement

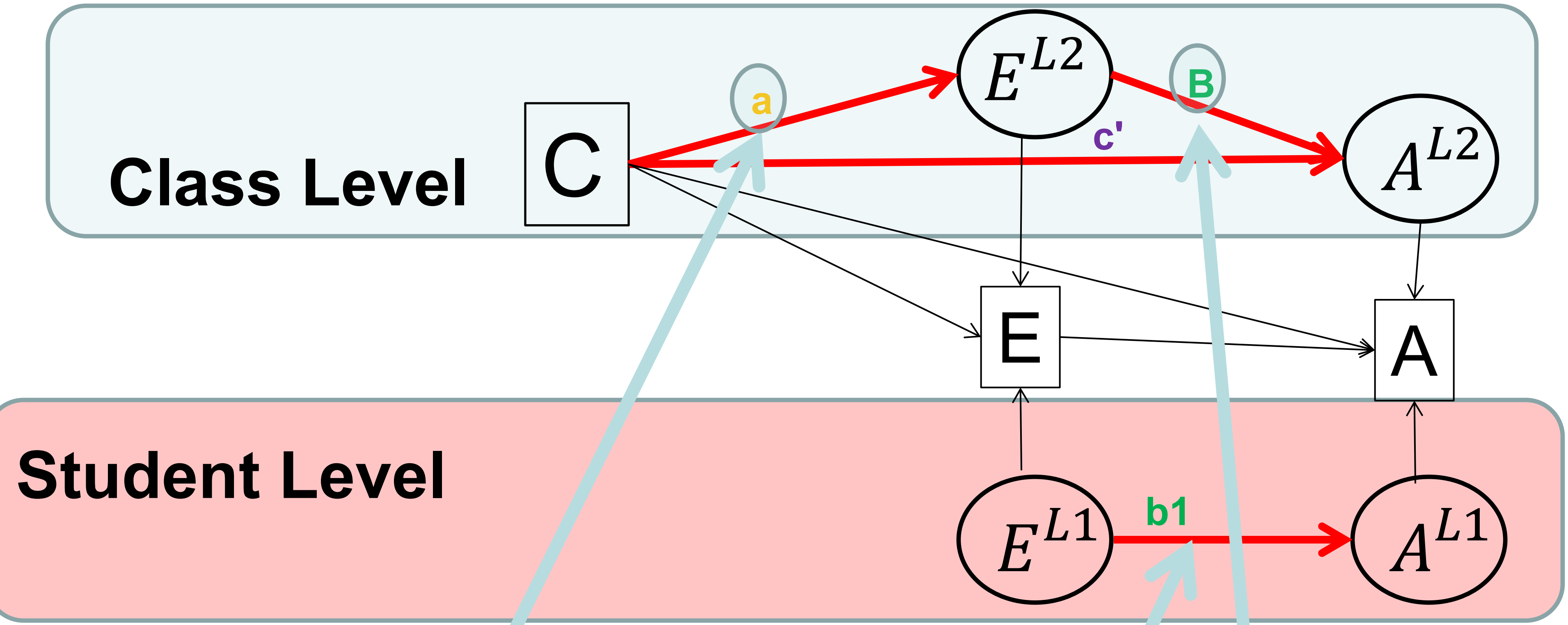
Graphical Illustration of 2-1-1 Mediation



Classroom Engagement



Classroom Engagement



Mediator Model

$$E_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + aC_j + u_{0j}$$

Outcome Model

$$Y_{ij} = \beta_{0j} + b_1 E^{L1} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + c'\ C_j + B\ E^{L2} + u_{0j}$$

Parameters for designing 2-1-1 mediation studies

Alpha: type 1 error rate (2 tailed): 0.05

a: treatment-mediator relationship effect size: 0.5

b1: mediator-outcome relationship at L1 effect size: 0.4

B: total mediator-outcome relationship effect size (B=b1+b2): 0.4

c': direct effect of treatment on outcome effect size: 0.1

rho2: Intraclass correlation for outcome: 0.2

rhom2: Intraclass correlation for mediator: 0.2

R21: outcome variance explained by covariates at L1: 0.5

R22: outcome variance explained by covariates at L2: 0.5

R2m1: mediator variance explained by covariates at L1: 0.5

R2m2: mediator variance explained by covariates at L2: 0.5

P: proportion of clusters in treatment: 0.5

n: L1 sample size: 20

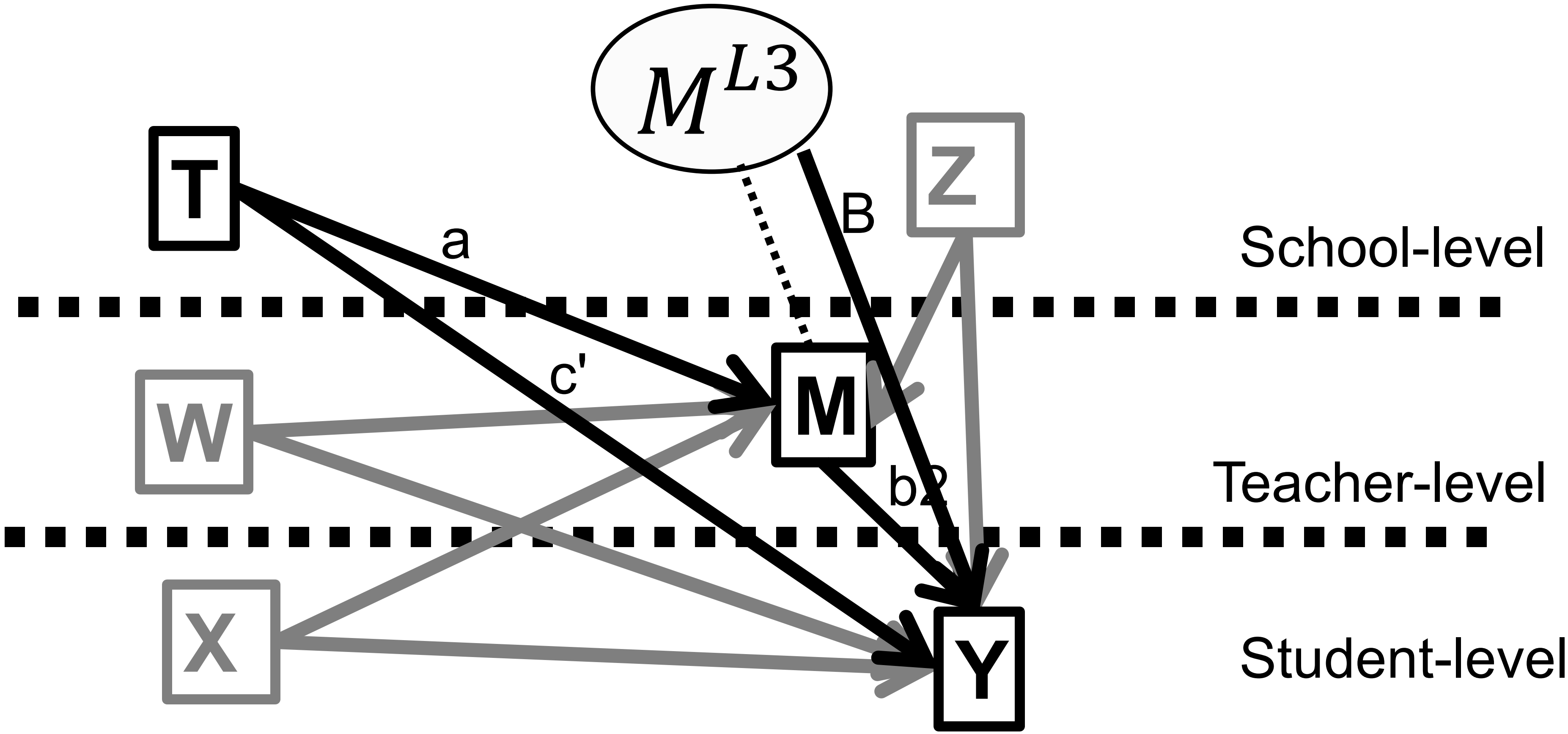
J: L2 sample size: ??

Power: 80%

3-2-1 Example

- School-randomized design
  - students nested within classrooms nested within schools
- Treatment: teacher professional development (assigned at school level)
- Outcome: students' achievement
- Mediator: teacher instruction
- Goal: 3-2-1 mediation
  - We are interested in designing a study to detect the extent to which the impact of participating in the PD program on student achievement is mediated by changes in instruction

Three-Level Example: 3-2-1



3-2-1 Parameters

$a = 0.50$ (treatment-mediator relationship [Cohen's d scale])

$B = 0.30$ (mediator-outcome relationship [Standardized regression scale])

$v_Y^2 = 0.10$ (unconditional outcome variance at school-level)

$\tau_Y^2 = 0.10$ (unconditional outcome variance at class-level)

$\sigma_Y^2 = 0.80$ (unconditional outcome variance at individual-level)

$\tau_M^2 = 0.20$ (unconditional outcome variance at school-level)

$\sigma_M^2 = 0.80$ (unconditional outcome variance at class-level)

$R^2_{Y\,L3} = R^2_{Y\,L2} = R^2_{Y\,L1} = 0.50$ (outcome variance explained at each level)

$R^2_{M\,L3} = R^2_{M\,L2} = 0.50$ (mediator variance explained at each level)

$P = 0.50$ (proportion of schools receiving treatment)

$n_2 = 4$ (classrooms/school)

$n_1 = 20$ (students/classroom)

nati

Additional Designs

- Three-level CRTs with mediator at any level
- Multisite individually-randomized designs with individual-level mediator
- Multisite cluster-randomized designs with cluster-level mediator

End of Session

- Break until 130pm

- Questions, Comments, & Feedback
  ben.kelcey@gmail.com

UNIVERSITY OF Cincinnati