More Than Main Effects: Designing Randomized Experiments to Test Moderators and Mediators of Student Achievement

James S. Kim
Harvard Graduate School of Education
kimja@gse.harvard.edu

STEM Education Research Designs: Conceptual and Practical Considerations for Planning Experimental Studies

November 2, 2009

Agenda

- What are moderators and mediators (conceptual framework)
- Concrete example from RCT of READ 180 (statistical framework)
  - Discuss results from RCT using different statistical models
  - Conclude with key lessons relevant to RCTs in various content areas (literacy, math, science)

I. Why should we study moderators and mediators?

- Policymakers and researchers
  - For whom does the intervention work best?
  - What is the causal mechanism that improves student achievement?

Baron & Kenny (1986): Moderator model

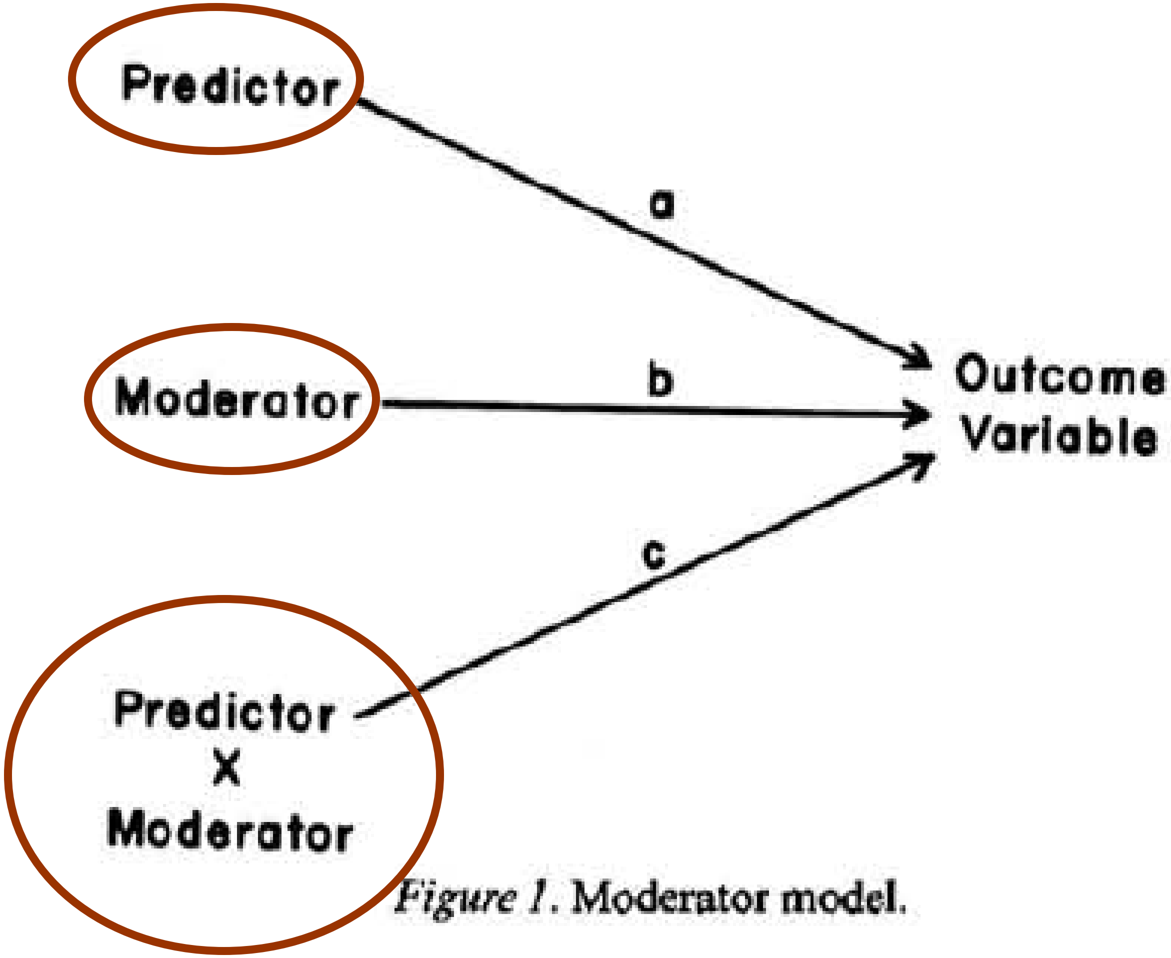


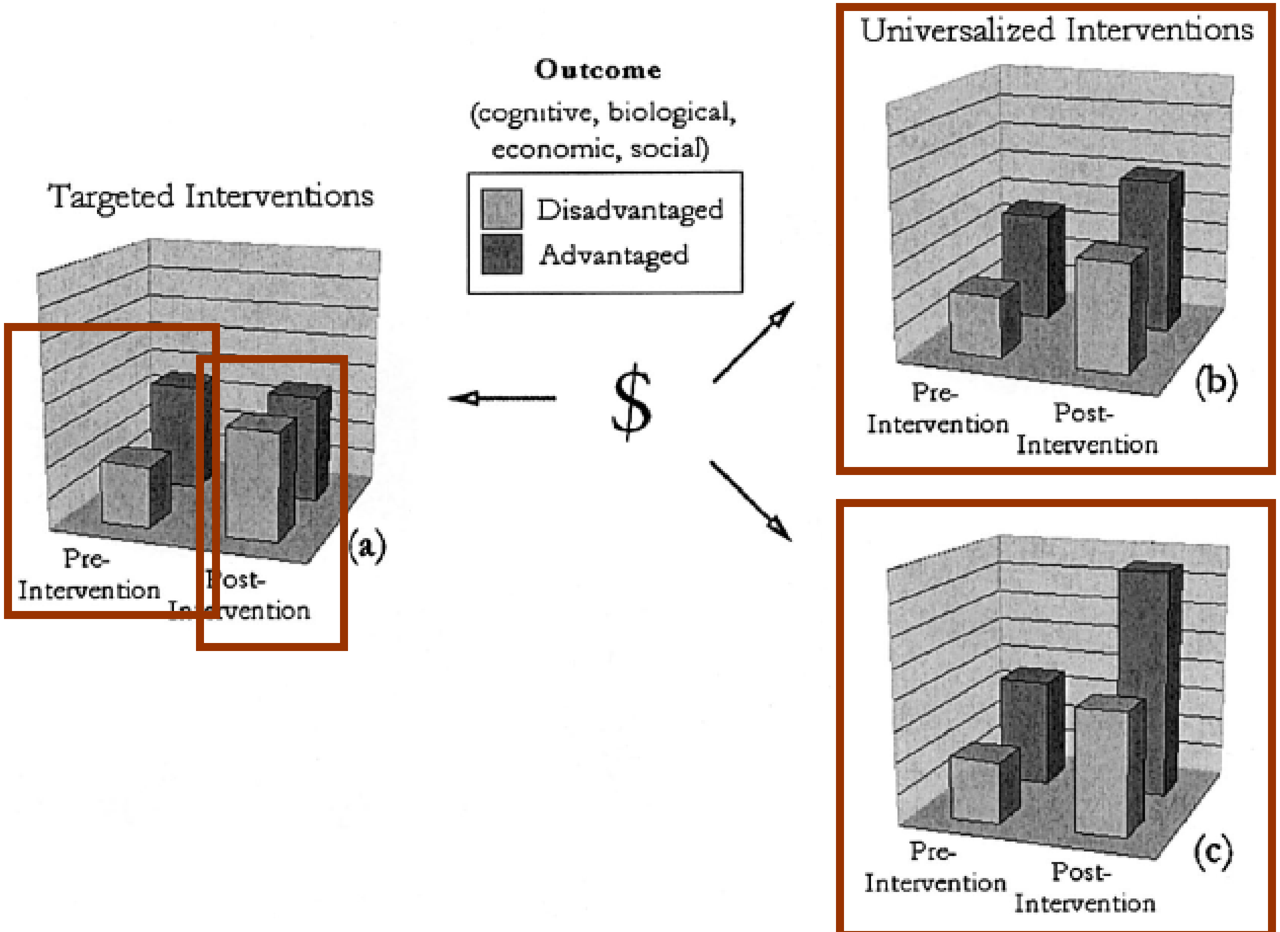
Figure 1. Moderator model.

"Moderator is a qualitative (e.g., sex, race, class) or quantitative variable (e.g., level of reward) that affects the direction and/or strength of the relation between an independent variable and a dependent or criterion variable" (p. 1174)

Moderators and achievement gaps

- Ceci & Papierno (2005) American Psychologist
  - “The rhetoric and reality of gap closing (When the “have-nots” gain but the “haves” gain even more)”

Figure 2
Hypothetical Representation of the Matthew Effect When Targeted Interventions Are Universalized



Note. (a) Disadvantaged children gain significantly—sometimes closing a preexisting gap entirely or at least a major portion of it—following a targeted intervention. (b) The preintervention gap remains following universalized intervention, with both increasing their preintervention scores by the same amount. (c) The preexisting gap is exacerbated because the advantaged children gain disproportionately from an intervention.

Performance-based benefits

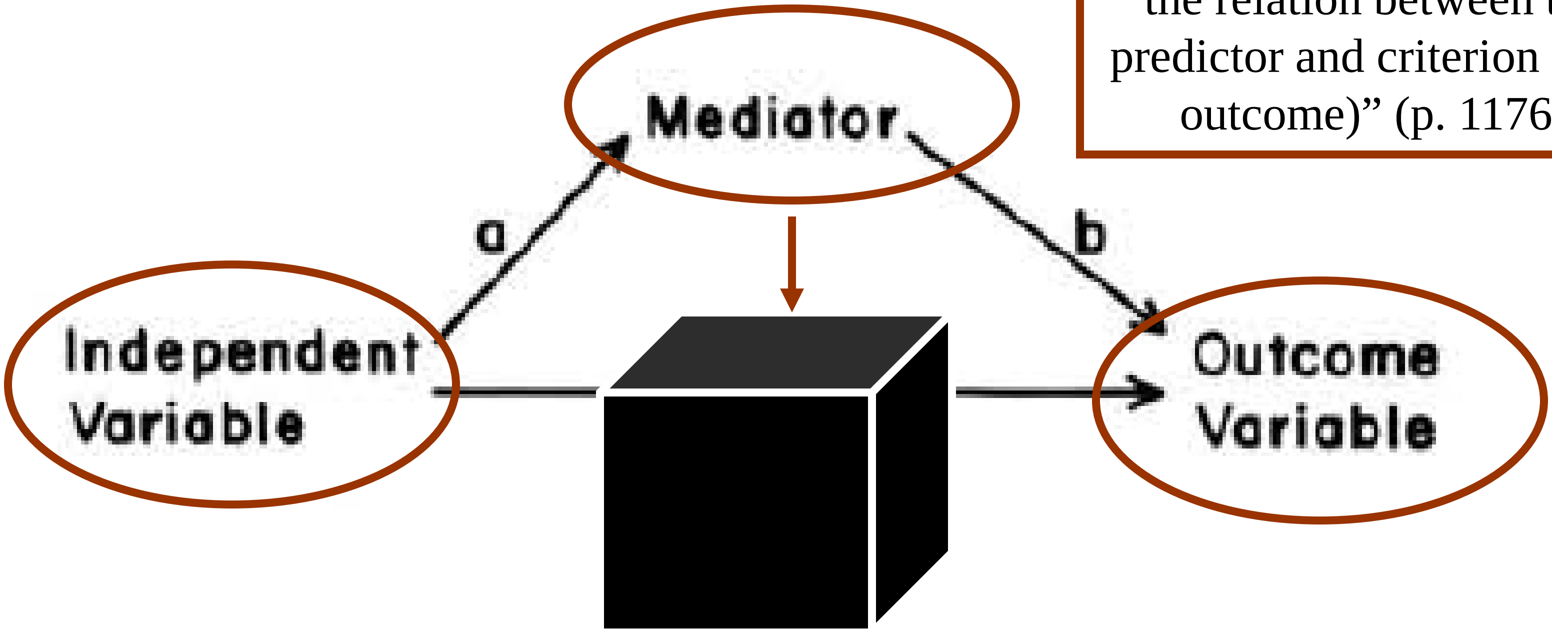
- Matthew effects in education
  - Rich get richer, poor get poorer
  - Hence, such interventions result in greater progress for the gifted children than for their nongifted counterparts even though the latter usually also make some progress (p. 153)
- Policy examples
  - Gifted > non-gifted in math program
  - Good readers > poor readers in print exposure, word reading ability, comprehension

Utilization-based benefits

- Universal policies can lead to unequal uptake
  - Public policies are often made universal to garner widespread support and may have different levels of use based on family SES of students
- Policy examples
  - GI Bill, Hope Scholarships favored students from middle- and upper-class homes

Baron & Kenny (1986): Mediator model

Mediators “account for the relation between the predictor and criterion (i.e., outcome)” (p. 1176)



Mediators of learning gains

- Intention-to-treat (ITT estimates)
  - Impact of being offered an intervention
  - Randomly assigned (and offered) to treatment
- Treatment-on-treated (TOT estimates)
  - Impact of actually receiving an intervention
  - Number of years child uses voucher to attend private school
  - Number of days child attends after-school reading and math program

II. RCTs of READ 180: Why evaluate a mixed-methods literacy intervention?

- By Grade 4, schools assume kids can use reading as a tool for learning content
  - Heterogeneity of reading difficulties in grade 4
  - Mixed-methods intervention that address multiple skill weaknesses (eclectic)
  - Additional instruction to accelerate learning of poor readers (supplemental)
  - Kim's KSA Method

Addressing the adolescent literacy crisis

- The literacy problem in the upper elementary, middle grades, and high school grades:
  - Struggling readers struggle for different reasons
  - A theory of reading can inform interventions to address the problem

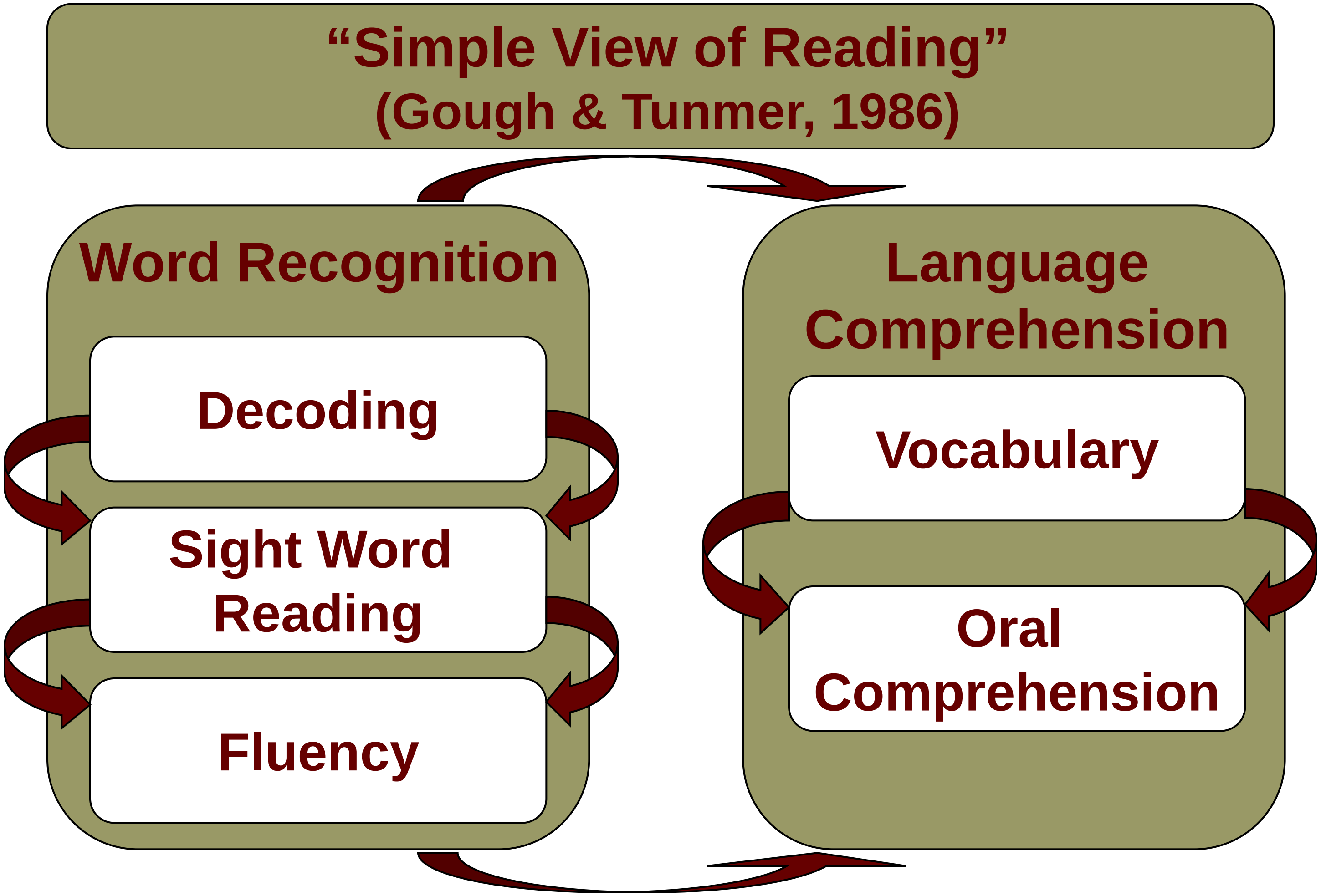
Struggling readers (all below proficient-NCLB) struggle for different reasons

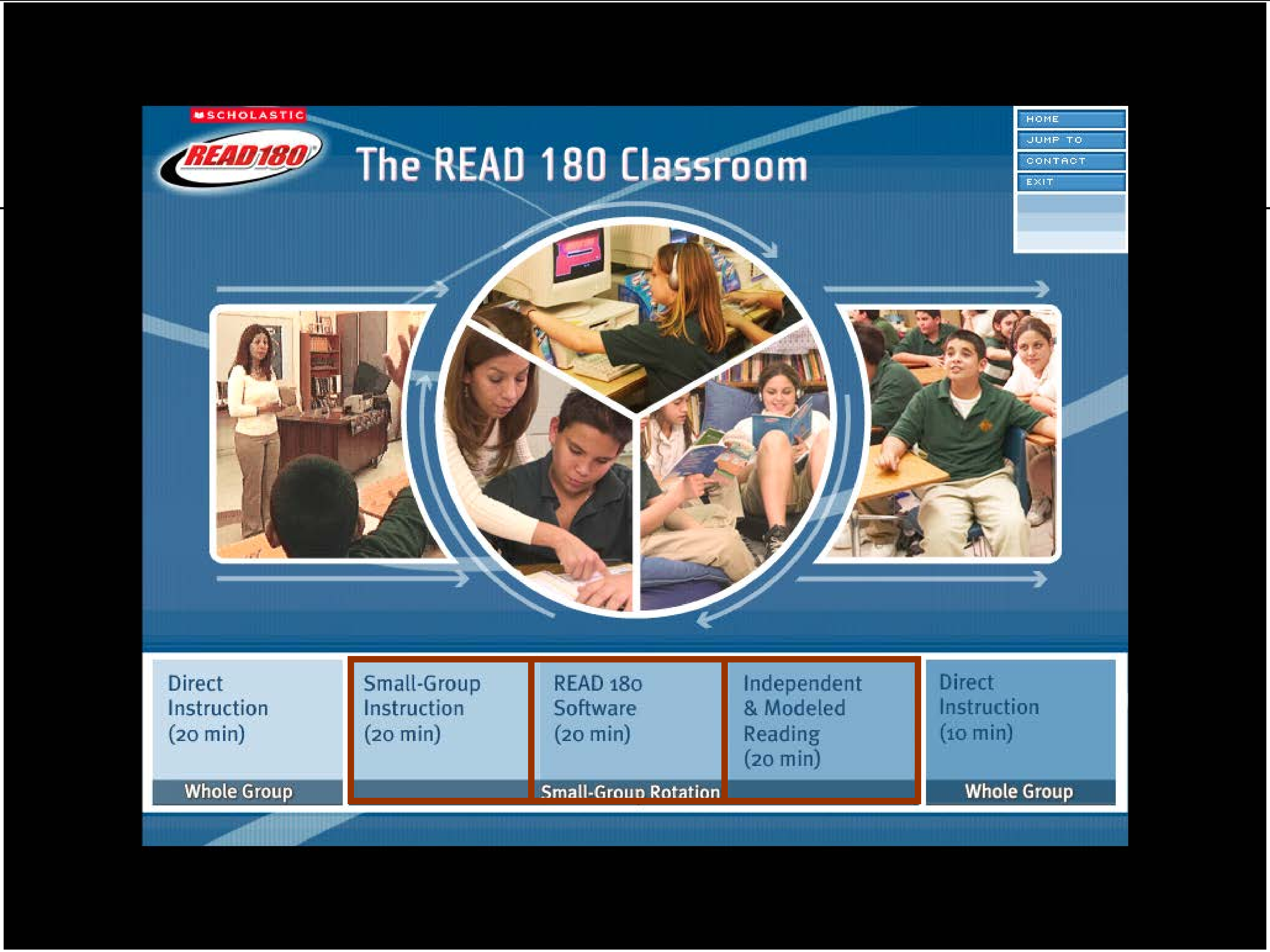
TABLE 4

Cluster Analysis

Cluster	*N*	Word identification	Meaning	Fluency	WASL score	% Not Eng.	% Low SES	Writing score	% L1	% L2
1	10	.37	–.58	.79	374	60	90	49	50	50
2	9	.64	–.33	.48	378	67	89	47	44	56
3	16	–.07	–.34	.72	380	56	81	53	31	69
4	19	–.42	.49	–.27	384	16	42	49	32	68
5	11	.09	.75	–.40	388	09	36	58	09	90
6	15	.37	.48	–.31	380	27	67	60	47	53
7	12	.38	–.34	–.09	377	58	67	57	50	50
8	6	.34	–.13	–.55	372	50	67	46	50	50
9	2	–2.55	–1.68	–.92	359	50	100	25	100	0
10	8	–1.07	–.31	–.44	374	12	75	38	63	37

A theory of reading can inform interventions to address the problem




SCHOLASTIC
READ 180
The READ 180 Classroom
HOME
JUMP TO
CONTACT
EXIT
Direct Instruction (20 min)
Whole Group
Small-Group Instruction (20 min)
READ 180 Software (20 min)
Independent & Modeled Reading (20 min)
Small-Group Rotation
Direct Instruction (10 min)
Whole Group

What do we know about efficacy

- Scholastic claims large impacts for lower-performing readers (2 grade levels behind)
- READ 180 evidence base
  - Slavin et al. (2008) review of quasi-experimental evaluations
  - Main effects on single measure (reading comprehension)
  - Focus on ITT (impact of offering intervention)
  - Few rigorous RCTs that examine moderators and mediators of student achievement on multiple measures

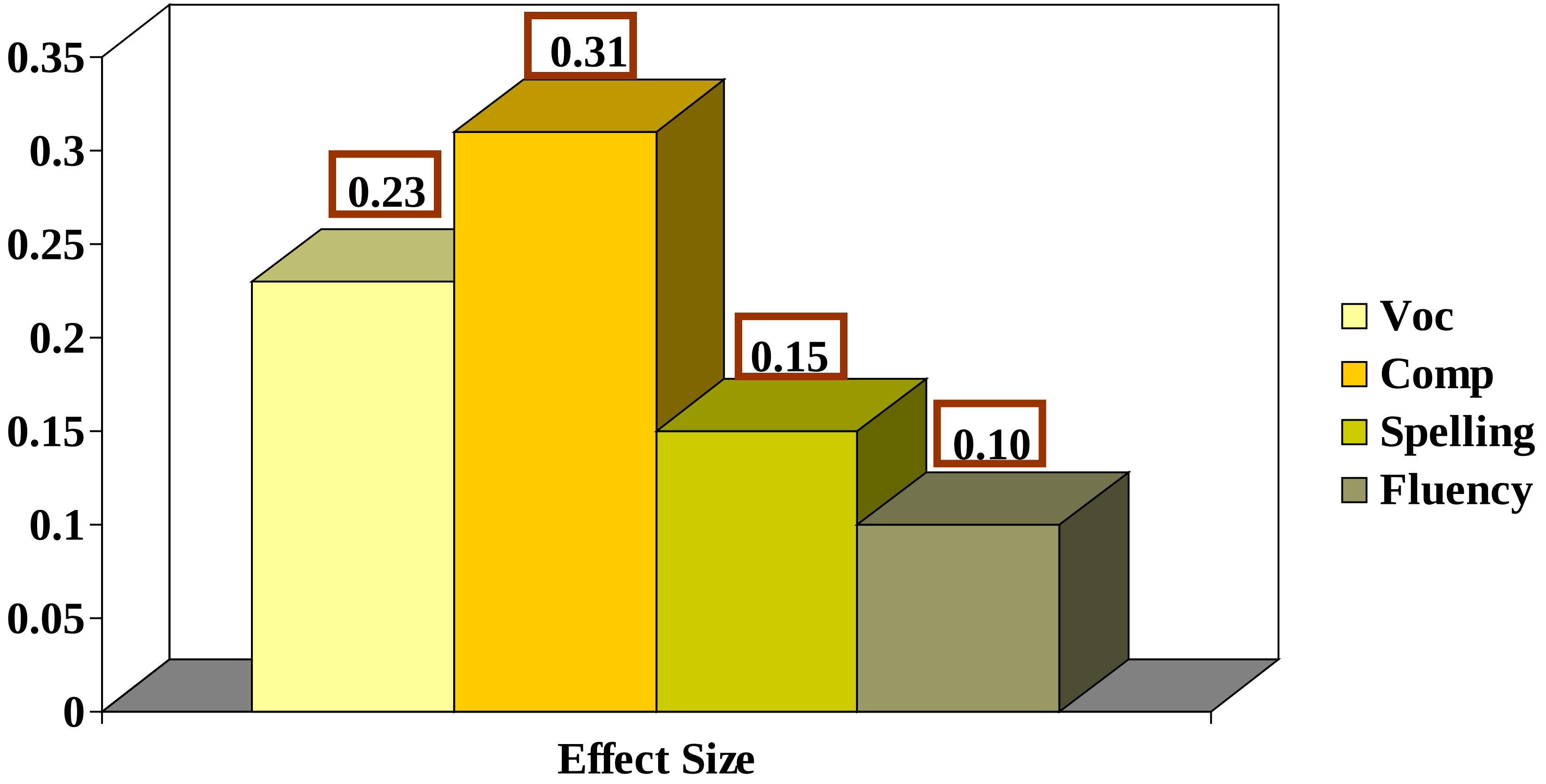
READ 180 RCT in After School

- Design
  - 312 students blocked by school and grade (4, 5, 6) and randomly assigned to R180 or district after school
  - Small-sized urban district in MA
- Sample
  - 66% Black and Hispanic students, 69% low-income
  - 95% below proficient on 2006 MCAS ELA
- Measures
  - Pretest covariate: DIBELS oral reading fluency
  - SAT 10 vocabulary, comprehension, spelling, fluency
  - Attendance

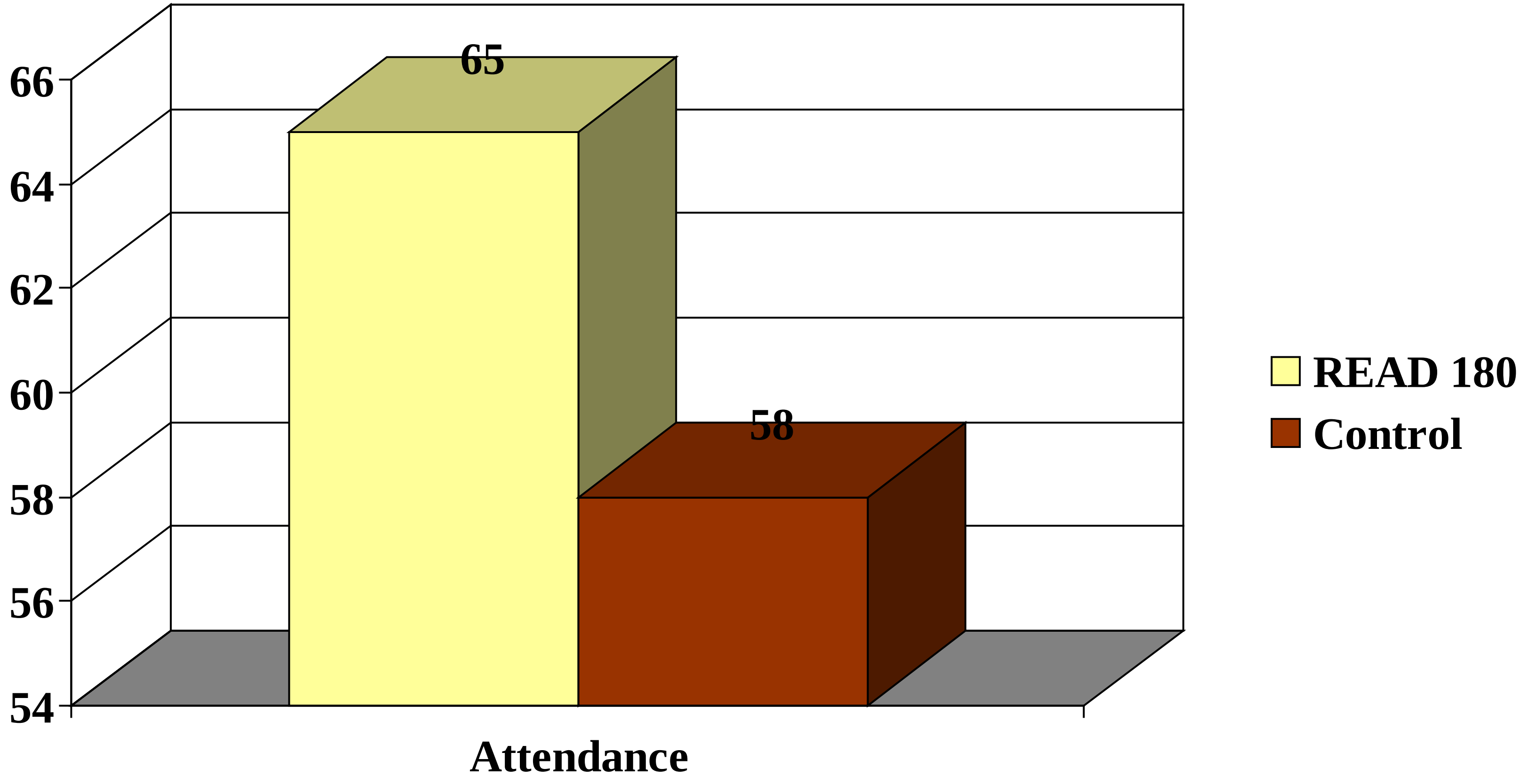
What's the main effect of R180?

$$\text{Model}: Y_i = \alpha + \boxed{\beta_i}(R180) + \gamma_i Z_i + \varepsilon_i$$

Effect sizes (SD) for reading outcomes



Attendance (ES = 7 days, .28SD)



Main effects inform policy & research

- Policy: If we implement the program, what impact will it have on achievement?
- Research: Is the theory of instruction underlying the intervention supported by experimental evidence?

What are the key moderators? (How many interactions are worth testing?)

- Think of our data / sample (NCLB subgroups, achievement levels, family variables, etc…)
- “Once we attend to interactions, we enter a hall of mirrors that extends to infinity. However far we carry our analysis—to third order or fifth order or any other—untested interactions of a still higher order can be envisioned.”
  - Cronbach (1975, p. 119)

Consider the theoretical mechanisms of the moderator variables

- Performance-based moderator variables
  - Kids above and below proficient (grade level) on MCAS English language arts
  - Prior skill level might moderate gains
- Utilization-based moderator variables
  - Younger kids have higher levels of motivation than older kids and may be more likely to actively attend after-school program
  - Grade level might moderate gains

Prior reading (DIBELS oral reading fluency) and motivation (ERAS)

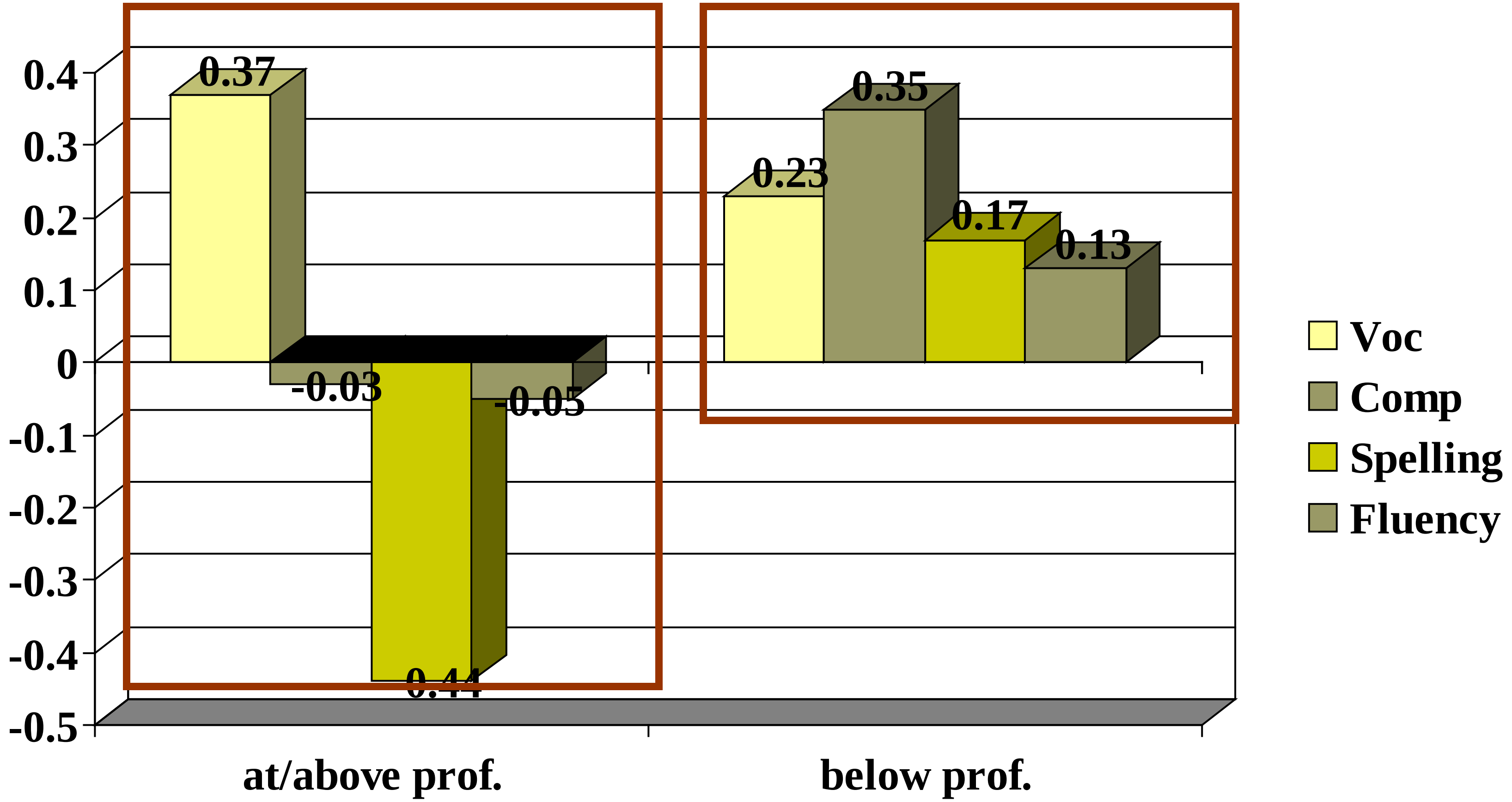
- Fluency (1 prompt)
  - Proficient = 101 words read in 1 minute
  - not proficient = 90 words read in 1 minute
  - .40 SD difference
- ERAS (20 item scale, nationally normed)
  - G4 = 45 PR (percentile)
  - G5 = 33
  - G6 = 37

Does prior achievement (proficient NCLB test) moderate impact of R180 on student achievement?

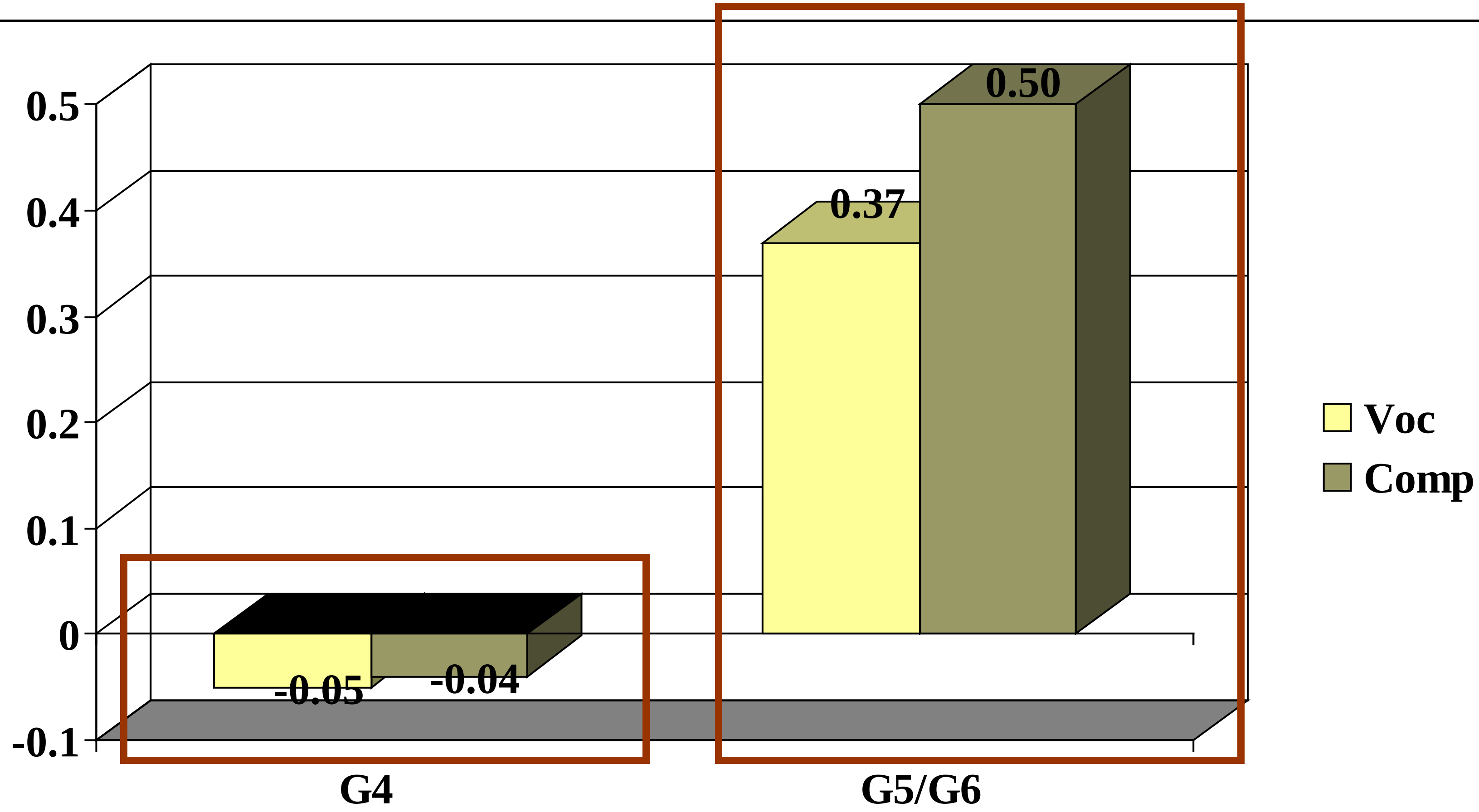
$$Y_i = \alpha + \beta_i (R180) + \gamma_i Z_i + \boxed{\delta_i} (R180 * Perf.MCAS) + \varepsilon_i$$

- The parameter δ indicates the interaction of R180 and prior reading level (above or below proficient)

Does prior achievement (proficient NCLB test) moderate impact of R180 on student achievement?



Effect Sizes by Grade Level



Moderators can elucidate mediators

- Why do effects on most outcomes appear larger for kids below proficient?
- Why are effects on vocabulary and comprehension larger for students in grade 5/6 than in grade 4?
- Mediator model: attendance is causal mechanism through which interventions improve student achievement

Treatment effects on attendance

Group	R180 (rounded)	Control (rounded)	Raw diff (days)	T (p)
Below prof	65	58	7	2.24 ($p < .05$)
At/above prof	75	72	3	.45 (n.s.)
G4	68	66	2	1.27 (n.s.)
G5/6	64	53	11	2.82 ($p < .01$)

How do we identify variables that mediate causal impact of intervention on student outcome?

- Why not regress posttest score on “days” of attendance? Won’t that give you a causal estimate of attendance?

$$\mathrm{OLS}: Y_i = \alpha + \boxed{\beta_1}(days)_i + \boxed{\gamma_1} X_i + \boxed{\varepsilon_i}$$

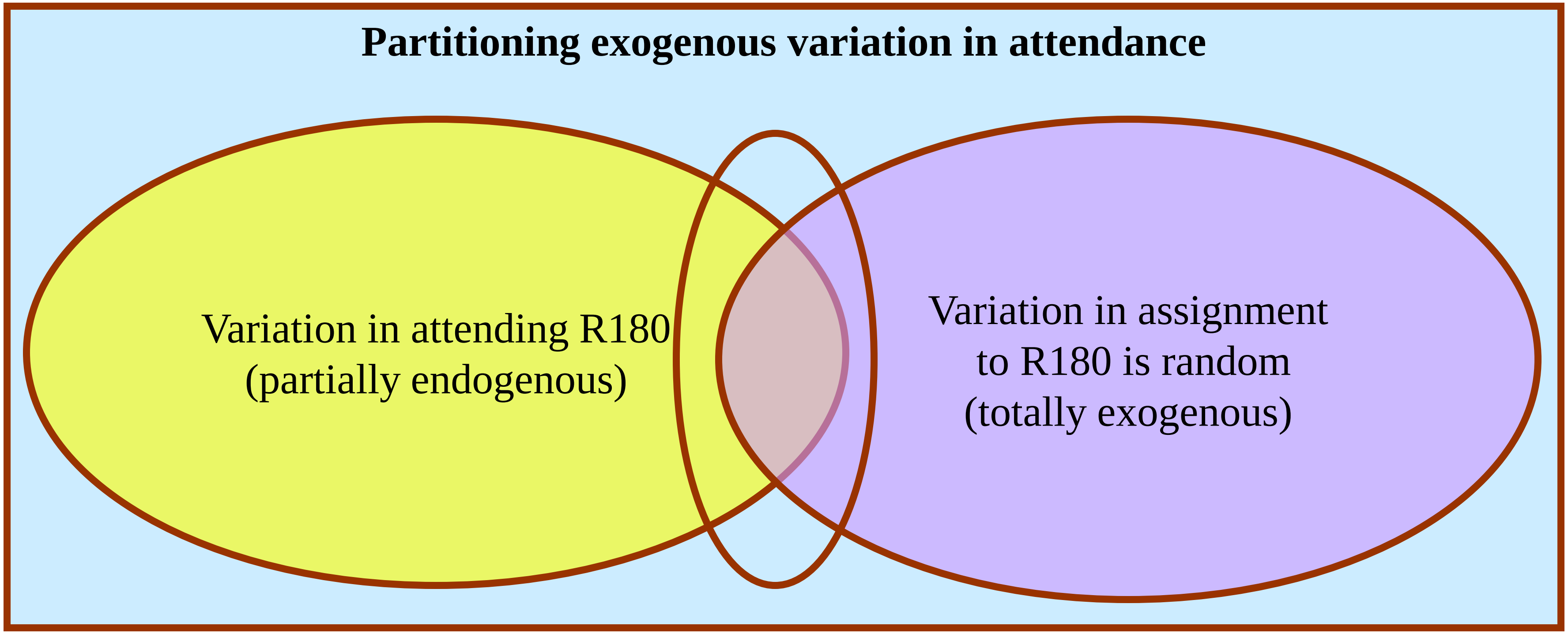
✓Days: Two sources of variation: (1) exogenous (variation outside kids, e.g., external research rolls dice) and (2) endogenous (variation inside kids, e.g., motivation, family factors)

✓“Days” yields biased estimates of R180: The coefficient for days includes “exogenous” variation (kids got randomly assigned to program) and “endogenous variation

✓Unbiased estimate of β assumes that correlation between β and ε = 0. This holds only if β captures ONLY exogenous variation in assignment.

Using instrumental variables to identify causal effects

- Attendance driven by (a) exogenous factors (i.e., random assignment) and (b) endogenous factors (i.e., motivation)



Defining an instrumental variable

- Valid instrumental variable
  - The instrument (random assignment variable) must be correlated with endogenous variable of interest (attendance)
  - The instrument must be uncorrelated with unobserved factors that affect attendance (i.e., instrument must be uncorrelated with residuals)
  - The instrument must affect posttest outcome ONLY through attendance (exclusion restriction)

Two stage least squares estimation

Stage 1 Model: Predict attendance using assignment status (randomized)

$$(1)\ \widehat{Attendance}_i = \beta_0 + \beta_1 R180_i + \beta_2 X_i + \mu_i$$

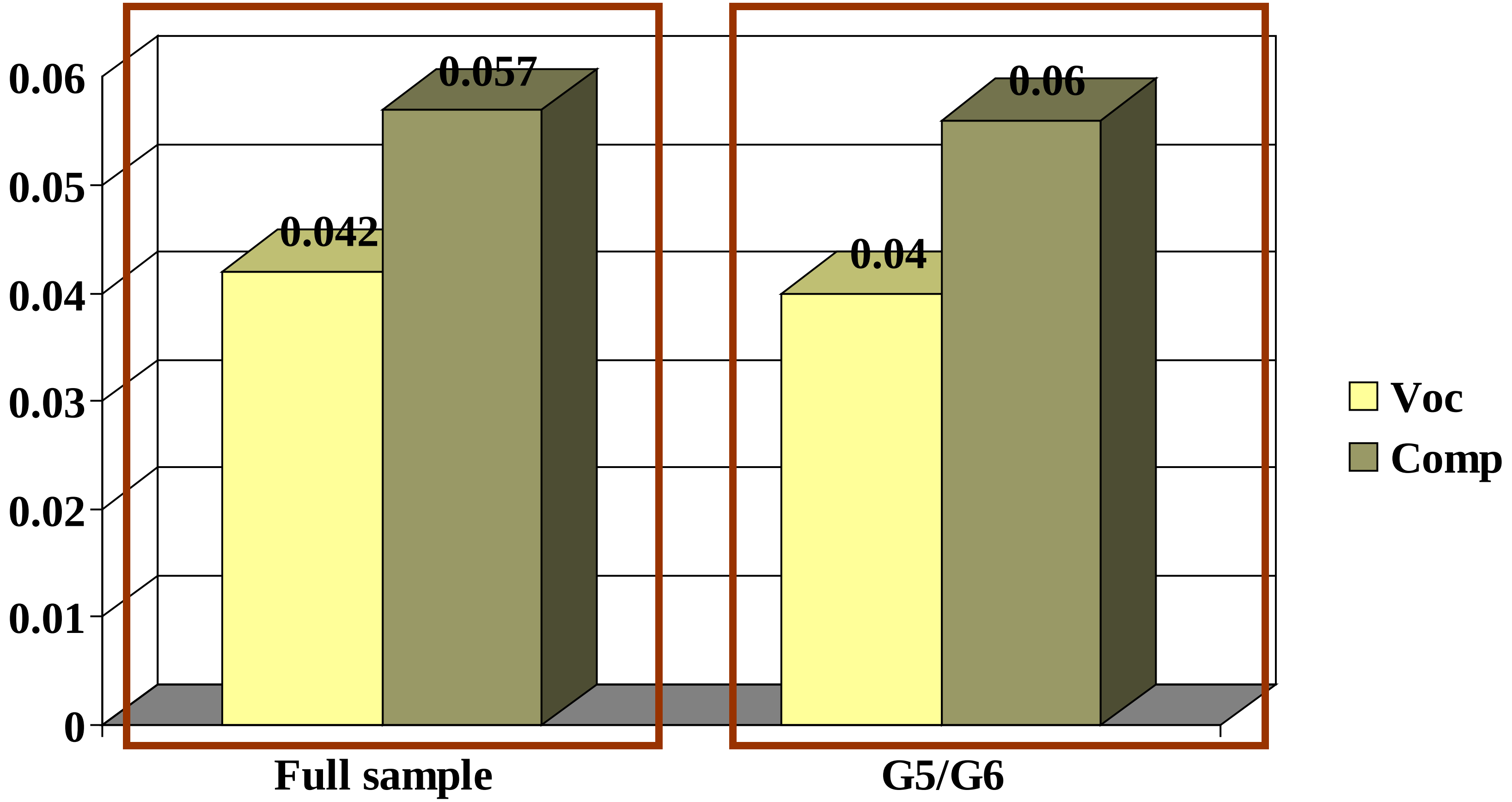
Stage 2 Model: Using exogenous variation in attendance to predict posttest score

$$(2)\ Y_i = \delta_0 + \delta_1 \widehat{Attendance}_i + \delta_2 X_i + \varepsilon_i$$

OLS v. IV estimates of attendance on vocabulary and comprehension

†$p < .10$ * $p < .05$	Vocabulary	Comprehension
Full sample (IV)	1.45 (.85) t = 1.70†	1.66 (.83) t = 1.98*
Full sample (OLS)	.011(.08) t = .14 (n.s.)	.112(.07) t = 1.63 (n.s.)
G5-6 (IV)	1.43(.68) t = 2.10*	1.71(.74) t = 2.31*
G5-6 (OLS)	.102(.096) t =1.05	.067(.087) t = .77

Impact of 1 additional day of attendance on posttest scores



Concluding lessons

- Lesson #1: RCTs allow us to test theories of instruction
- Lesson #2: Performance-based and utilization-based moderators of achievement can help us test substantively important interaction effects in RCTs
- Lesson #3: Moderators can elucidate mediators
- Lesson #4: Instrumental variables can extract causal impacts of attendance on student outcomes if key assumptions are met
- Lesson #5: What is the impact of the treatment on the treated (i.e., kids who attend and participate in an intervention)? Answers may depend on our statistical models
- THANKS!