TUTORING AND SPILLOVER EFFECTS:

Measuring Spillover Effects of High Impact Tutoring onto Peers of Low Performing Students

Eli Mogel - UCSD

HIGH IMPACT TUTORING: AN OVERVIEW

Definition:

- *One-to-One or Small Group Sessions*
- *≥ 3 Days a Week*
- *≥50 Cumulative Hours*

Policy

- Cost Effective: Effect-size/$1,000 (Harris, 2009)
  - Tutoring: 0.9SD
  - Computer Assisted Instruction: 0.6SD
  - Increased instructional time: 0.4SD
  - Early Childhood: 0.17SD
  - School Reform: 0.116SD
  - Reduced Class Size: 0.08SD
- Comparatively better at targeting low SES and performing students (Dietrichson et al. 2017)
- Scalable (Oreopolous et al. 2020)
- Make a policy related thing to understand

Research

- Unknown: Spill-over onto peers
- Implications for previous research and experimental designs:
  - Bias?
  - Peer Effects?
  - True Impact of Tutoring?

RESEARCH QUESTION

What effect does tutoring low performing students have on peers of the tutored students?

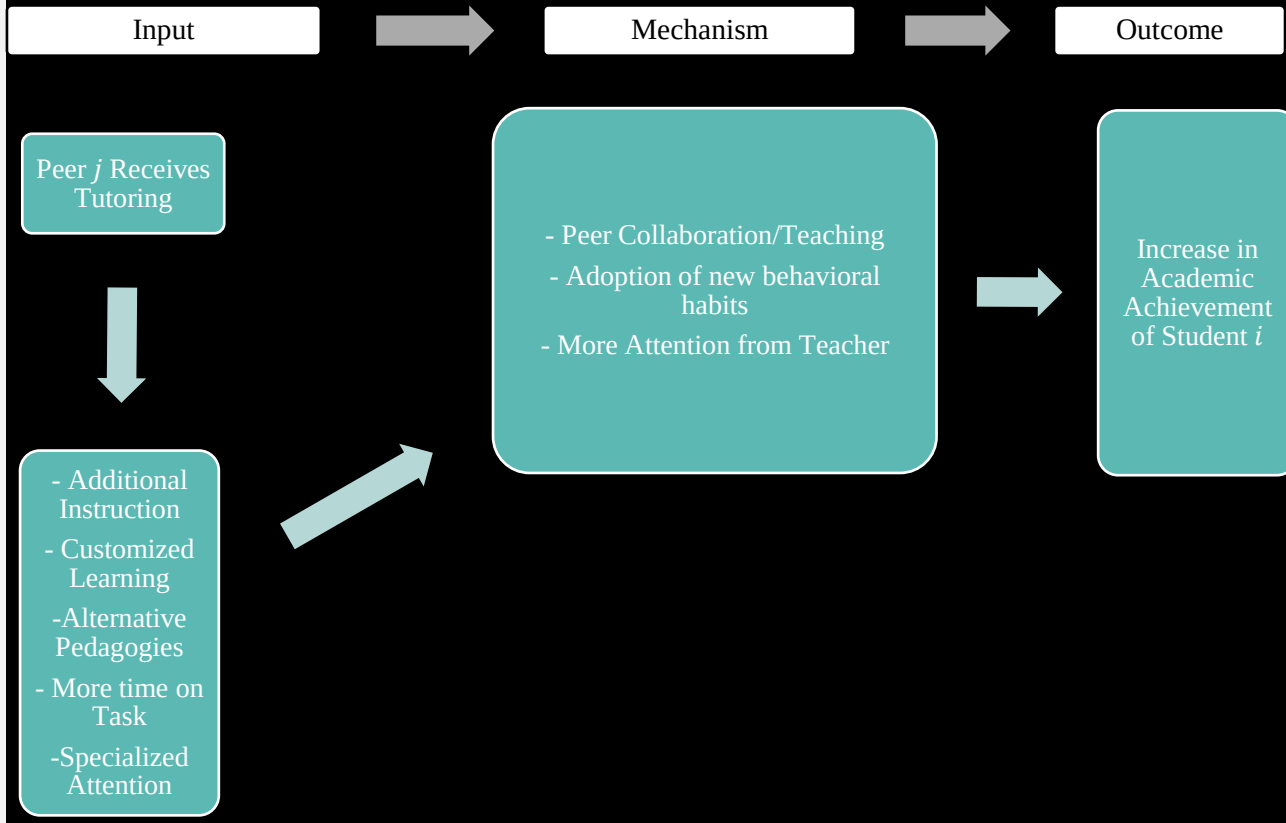
- Importance
  - Policy: Better understand effectiveness of program
    - Federal: Americorps, Department of Education
    - State/local
    - NGO
  - Research/Theory:
    - Tutoring literature: Fills gap that previous experiments were unable to.
      - Previous Designs unsuited for spillovers.
    - Peer literature: Studies how intervention effects existing peer groups
      - Instead of optimizing peer groups for most effective results.

Theory & Mechanisms:

Peer Effects: "An externality in which peers' background, current behavior, or outcomes affect an outcome of another individual"

- $Y_{ict} = f(X_{ict}, \alpha_c, \boldsymbol{P_{ict}})$
  - Y_{ict}: Academic or Behavioral Outcomes of Student i in classroom c in time t
  - X_{ict}: Characteristics of student i
  - α_c: Classroom characteristics
  - $\boldsymbol{P_{ict}}$: The peers of $student\ i$

- How could P_{ict} affect Y_{ict}?
  - Bad Apple/Shining Light: Provide poor/great academic or behavioral habits for student i to follows
  - Does P_{ict} relationship differentially impact Y_{ict}?
  - What if P_{ict} receives tutoring, what happens to Y_{ict}?



LITERATURE REVIEW

Tutoring

- Oreopoulos Et al. 2020
  - Pooled effect 0.37SD
    - Paraprofessional Tutors: 0.4SD
      - Grades 1 - 5: 0.41SD
      - 1-to-1 sessions: 0.46SD
      - 1-to-3 sessions: 0.38SD
      - During School: 0.48SD
      - After School: 0.2SD (Weakly Signif.)
- Large scale experiments
  - 15 studies to date with samples 400+
  - Gersten et al. (2015):
    - 4 State, 76 school math intervention: 0.34SD
  - Parker et al. (2019):
    - Cross grade 4-8 math intervention: 0.24SD

Peer Effects in Education

- Generally, better peers → higher achievement for student *i*
- **Note, read outcome as : For every 1.0SD↑ in peer math achievement, student *i* has a ***↑ in math achievement**
  - Exploited microdata & fixed effects, noisy but significant increases (Hanushek 2003, Beets 2004, Sass 2008): 0.03SD – 1.19SD ↑
  - Quasi-Experimental Designs:
    - Randomized Classroom Assignment (Kang 2007, Wang 2010): 0.27SD – 0.5SD ↕
    - Randomized Peer Formation, College Roommates (Sacerdote 2001, Foster 2006, Zimmerman 2003): Null – 0.06GPA
  - Experimental Design:
    - Financial incentive for female students → 0.16SD ↑ for male classmates (Kremer et al)
    - Randomized low achieving students into high achieving peer groups → 0.06SD↓ for low achieving students, segregation between ability in peer groups. (Carrell et al 2013).
- Lower performing students could receive strongest peer effects
  - Kang 2007: Bottom 25th percentile received 0.47 SD ↑ (74% higher than pooled estimate)
  - Other studies find likewise in vary size (Sass 2008, Imberman et al 2008).
- Peer formation helps determine achievement
  - Students seek relationships with academically similar individuals → these relationships can influence behavioral and academic choices in schools (Gremmen et al 2017, Shin 2018).

INTERVENTION

- Partner with National Student Support Accelerator to provide high intensity math tutoring to underperforming students in grades 3-5 in order to study spillovers onto peers/friends of non-tutored students.
  - Research Center at Brown University, devoted to "translating promising research about how tutoring can benefit students into action on the ground."
  - National Student Support Accelerator already has relationships with school districts and tutoring organizations in over 8 States:
    - California: San Jose Unified School District, Santa Barbara Unified School District
    - Georgia : Gwinnett County Public Schools
    - North Carolina : Guilford County Public Schools
    - Texas : Spring Independent School District, Pharr-San Juan-Alamo Independent School District
- Why Math?
  - Previous literature provides largest, less noisy outcomes than English/literature interventions
  - Most school districts above already have English/Literature tutoring interventions → Lower two-sided compliance problems
- Why Grades 3-5?
  - Elementary schools have highest treatment effect from tutoring
  - Administrative test score data is available starting in grade 2 for these states, allows for baseline measures

Funding Partners of NSSA

BILL & MELINDA GATES foundation

AS OUR PARENTS PLANTED FOR US, SO WILL WE PLANT FOR OUR CHILDREN

CHARLES AND LYNN SCHUSTERMAN FAMILY PHILANTHROPIES

WALTON FAMILY FOUNDATION

DESIGN: OUTCOME & MEASUREMENTS

- Data:
  - Administrative data from participating school districts
    - Previous year's math score
    - Demographic information of students
    - Current year's classroom assignment
    - Behavioral/academic punishments (if any)
  - Survey-Data
    - Friendship sheets:
      - Information that elicits 3 "best" friends in class
    - Why?
      - Allows us to study if spillover's effect size is related to relation of tutored student
- Study Sample: "Eligible Students"
  - Lowest 25% of students based on math score in previous year in current class.
    - Students that were present in district previous year
    - Students in "regular" classes.
  - Why?
    - Meaningful policy-wise
    - Cost effective use with available data
- Treatment: 20 week, 3-to-1, 50-minute, 3 days a week, tutoring sessions in math.
  - Taught by: Paraprofessionals already contracted with school districts, as part of National Student Support Accelerator.
  - Why?
    - Emulating other large scale RCTS that had 0.31 – 0.34 SD treatment effects
- Design: Saturated Treatment
  - Treatment 1: treat 50% of eligible students
  - Treatment 2: treat all but one eligible student
  - Control: business as usual
  - Why?
    - Allows for meaningful spillover effects
    - Efficiency: Treat 50% of 2 classes or 100% of 1 class?

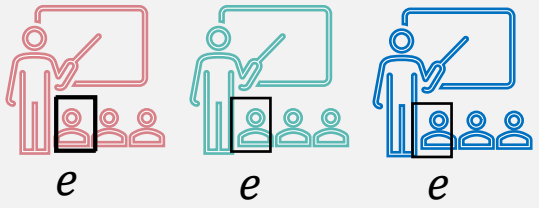
- Main outcome: Value Added Math Score on…
  - Spillover of tutoring onto nontutored eligible students.
  - Spillover of tutoring on eligible or noneligible students on friendship sheets.
- Secondary outcome:
  - Behavioral violations (if data is available) of nontutored eligible students or peers on friendship sheets of tutored students
  - Value added scores disaggregated…
    - by grade level
    - by sex
    - By sex of tutored/nontutored friend (MtM, FtF, FtM, MtF)

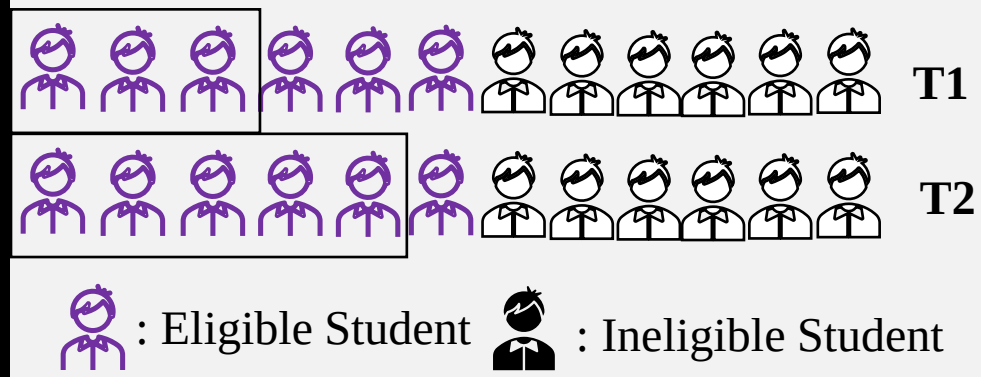
DESIGN: RANDOMIZATION

T1: 50% of eligible students are offered tutoring

T2: All but one eligible student is offered tutoring

C: Business as usual

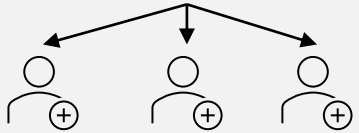




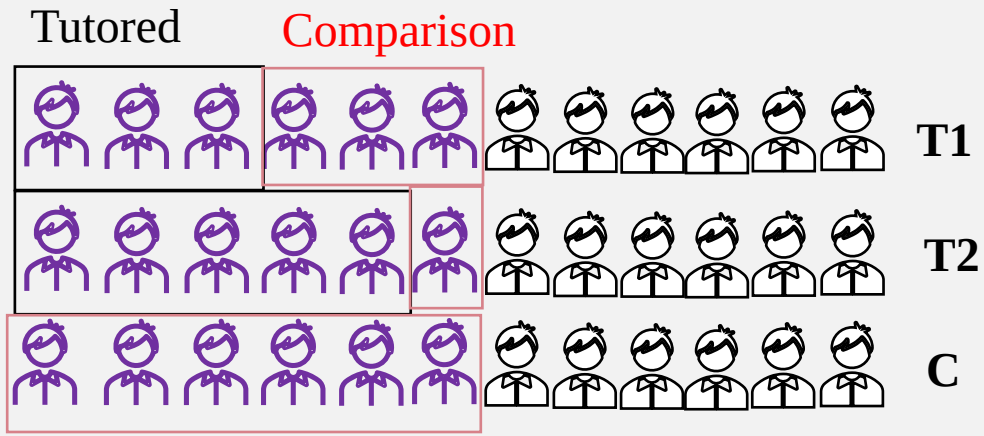
1. Offer high intensity math tutoring program to all eligible elementary schools in sample during year before intervention. Notify that conditional on acceptance, only a portion will randomly receive a yearlong tutoring intervention.
2. Of those that accept, identify average math achievement of grades 3-5, stratify on school district and math achievement to make triplets.
3. Randomly assign schools in stratification to either T1, T2, or Control
4. During **week 1** identify lowest 25% of students based on previous year math achievement on state exam in every classroom. These are the eligible students.
5. Within T1 & T2, randomly assign one classroom in grades 3, 4, and 5 to receive tutoring .
6. At **week 5**, offer tutoring program to parents of eligible students in T1, T2 classrooms. Parents that accept have their children randomized into treatment.

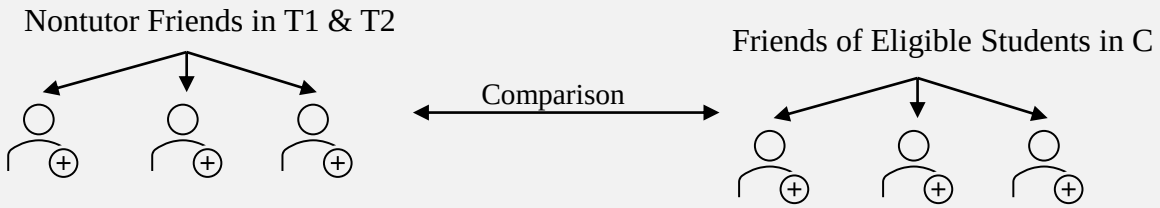
DESIGN: RANDOMIZATION

*"Who do you work and talk with the most in class?"**



**Not final language on friendship sheet*





7. At **end of week 6**, distribute friendship sheets for students to fill out.

8. At **week 7**, begin tutoring sessions with T1, T2 eligible students. Sessions are held during non-core classes (PE, Art, etc…). Sessions are done 3-to-1.

9. Sessions **continue for 20 active weeks of school**. Students take their state's standardized test and school year ends.

- Additional:
  - Comparisons will be:
    - Between lowest 25% of students in T1, T2, and C classrooms.
    - Friends on friend sheets of T1, T2, and C tutored students
  - T1, T2 schools provide one classroom per grade as to minimize interference (if we treat 2 3rd grade classrooms, there could be additional unrecorded spillovers).
  - All 3rd , 4th , 5th grade classrooms in C school can be used as controls (no fear of interference).

POWER CALCULATIONS

- What is an appropriate effect size?
  - Previous literature: 1.00SD ↑ in peers' achievement → ~0.3SD ↑ for student *i*
  - Large scale tutoring: 0.3SD ↑ for tutored's achievement → **~ 0.09SD ↑ for nontutored achievement**
  - Possible T1, T2, C classrooms
    - 240 students grades 3,4,5 in each school (Texas school data)
    - 25 students in each classroom (National avg classroom size)
    - $\frac{240}{25} = \sim 9$ classrooms in grades 3,4,5 per school
    - **T1 Schools: 3 classrooms in study**
    - **T2 Schools: 3 classrooms in study**
    - **C Schools: 9 classrooms in study**
  - Possible tutored students
    - 25 students in each classroom; 25% = 6 eligible students
    - **T1: 50% of eligible students = 3 tutored students**
    - **T2: All but one eligible student = 5 tutored students**
    - **C: Business as usual**

Fixed Parameters

- N = 6 eligible students
- ICC = .11 (Hedges, 2007)*
- R^2 = .5 (Hedges, 2007)*
- Control = 50% of schools
- T1 = 30% of schools
- T2 = 20% of schools
  - *ICC, R^2 taken from Education based RCT meta-analysis

Schools	75% Compliance	50% Compliance
350	SNT: 0.089SD; Slope-SNT: 0.19SD	SNT: 0.12SD; Slope-SNT: 0.23SD
400	SNT: 0.08SD; Slope-SNT: 0.18SD	SNT: 0.1SD; Slope-SNT: 0.22SD
450	SNT: 0.07SD; Slope-SNT: 0.17SD	SNT: 0.96SD, Slope-SNT: 0.21SD

Power Calculation & MDES

- ***Need at least 260 schools to comply for MDES of 0.09SD***

COSTS OF EXPERIMENT

- Costs mainly to fund at-cost tutoring for 20 weeks:
  - Tutors:
    - 20$ per – hour
    - 60$ per student operational fee (supplies, etc.)
    - 2.5 hours per week
    - 20 weeks
    - 3 to 1 sessions
    - 20$ * 2.5 hours * 20 weeks + 60$* 3 students = **1180$ per 3 students**
  - T1:
    - ~9 Students need tutoring (3 per grade)
    - 1180$ * 3 groups * 78 Schools = **$276,120**
  - T2:
    - ~15 Students need tutoring (5 per grade)
    - 1180$* ~ 5 groups * 52 Schools * 2 tutors = **$613,600**
- Research based costs:
  - Research manager, two-year salary: **$145,000**
  - Research assistants, two-year salary: **$100,000**
  - Total Costs:
    - **$276,120+ $ 613,600 + $145,000 + $100,000 = $1,134,720**

MODEL AND ISSUES

Main Model: $Y_{ic} = \pi_0 + \pi_1 Z_{ic} I_{ic} + \pi_2 I_{ic} + \pi_3 Y_{ic,-1} + X_i + B_{ic} + G_c + S_{ic} + e_{ic}$

- Y_{ic} = End-line Math Score of Student I in class C
- Z_{ic}= Indicator if tutored student in treatment classroom(TE)
- I_{ic} = Indicator if in eligible student in treatment classroom(SNT)
- $Y_{ic,-1}$ = baseline math score
- X_i = baseline covariate information
- B_{ic} = blocking fixed effect
- S_{ic} = School fixed effect
- G_c = grade fixed effect

- Issues
  - Compliance:
    - Schools that don't win tutoring might be less responsive in carrying out surveys.
    - Schools might already have tutoring programs in math, outside tutoring.
  - Interference:
    - Cross-grade spillover within T1, T2 schools?
  - Attrition:
    - Differential attrition from study? Control students leave district in response to no tutoring, treated students leave study due to displeasure of tutoring.
    - Small sample sizes, attrition can harm measurements
  - Further Considerations:
    - Crafting friendship sheet that adequately answer question across grade levels
    - Differential impact of tutoring by school district/tutoring organization

MORE

- Saturation Slope Model: $Y_{ic} = \pi_0 + \pi_1(50\%)Z_{ic} + \pi_2(100\% - 1)Z_{ic} + \pi_3(50\%)I_{ic} + (100\% - 1)I_{ic} + \pi_3 Y_{ic-1} + X_i + B_{ic} + G_{ic} + S_{ic} + e_{ic}$
  - Same as before but have a 100% and 50% saturation variables

- Friendship Sheet Model: $Y_{ic} = \pi_0 + \pi_1 Z_{ic} + \pi_2 I_{ic} * Fr_{ic} + \pi_3 Fr_{ic} + \pi_4 Fr_{ic} + \pi_5 Y_{ic-1} + X_i + B_{ic} + G_{ic} + S_{ic} + e_{ic}$
  - Same as before but Fr_{ic} is indicator if you are on the friend sheet. Therefore, the interaction is the heterogenous effect of being on the friend sheet and also being an eligible student
  - Fr_{ic} alone is being an ineligible friend