How Does Your Kindergarten Classroom Affect Your Earnings?
Evidence from Project STAR

Raj Chetty, Harvard
John N. Friedman, Harvard
Nathaniel Hilger, Harvard
Emmanuel Saez, UC Berkeley
Diane Schanzenbach, Northwestern
Danny Yagan, Harvard

July 2010
Motivation

• Large literature on how interventions such as better teacher quality or smaller classes affect achievement as measured by test scores

• Much less evidence on whether interventions that increase test scores improve outcomes such as earnings

• Problem: few datasets link information on early childhood test scores with data on adult outcomes

  • We link data from the STAR experiment to US tax records to evaluate the long-term impacts of education interventions

→ Question: Are higher test scores a good proxy for improvements in adult outcomes?

  • Do small classes and better teachers/peers improve adult outcomes to the extent they improve test scores?
Project STAR: Background

- Student/Teacher Achievement Ratio (STAR) experiment is one of the most widely studied education interventions
  - Conducted from 1985 to 1989 in Tennessee
  - One cohort of 11,571 children in grades K-3 at 79 schools
  - Most children in the experiment born in 1979-80 → graduate high school in 1998
  - Students and teachers randomized into classrooms within schools
  - Class size differs: small (~15 students) or large (~22 students)
  - Classes also differ in teachers and peers
  - Only one cohort treated → no repeat teacher observations
Project STAR: Background

- Large literature has investigated impacts of STAR on test scores
  - Students in small classes have 5 percentile point (0.2 sd) higher test scores in K-3 (Krueger 1999)
- Test score gains fade out to 1-2 percentiles by grade 8
  - Similar fade out effects observed in other early childhood interventions (e.g. Currie and Thomas 1995, Deming 2009)

→ Concern that early test score gains may not translate into impacts on adult outcomes
United States Tax Data

- Access to selected variables in anonymous U.S. tax records to conduct research on behavioral responses to economic policies
- Dataset covers full U.S. population from 1996-2008
- Approximately 90% of working age adults file tax returns
- Third-party reports yield data on many outcomes even for non-filers
  - Employer and wage information from W-2 forms
  - College attendance information for all kids from 1098-T forms
- 93.4% of STAR records were linked to tax data
- Match rate orthogonal to treatments
<table>
<thead>
<tr>
<th></th>
<th>Mean (1)</th>
<th>St. Dev. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Wage Earnings (2005-07)</td>
<td>$14,790</td>
<td>$13,808</td>
</tr>
<tr>
<td>Fraction With Zero Earnings ('05-'07)</td>
<td>13.6%</td>
<td>34.3%</td>
</tr>
<tr>
<td>Attended College in 2000 (age 20)</td>
<td>26.5%</td>
<td>44.1%</td>
</tr>
<tr>
<td>Ever Attended College by age 28</td>
<td>46.7%</td>
<td>49.9%</td>
</tr>
<tr>
<td>Parental Household Income ('00-'07)</td>
<td>$39,030</td>
<td>$37,953</td>
</tr>
<tr>
<td>Fraction Black</td>
<td>36.4%</td>
<td>48.1%</td>
</tr>
</tbody>
</table>
Outline

1. Test scores and adult outcomes in the cross-section
2. Re-evaluate validity of STAR experimental design
3. Class size impacts on adult outcomes
4. Teacher/peer impacts on adult outcomes
5. Conclusion: Cost-Benefit Analysis
Part 1: Cross-Sectional Correlations

- Begin with OLS estimates of return to higher early childhood test scores by correlating KG test scores with adult outcomes.

- Useful to benchmark estimates obtained from randomized interventions.

- Test score: Percentile score on Stanford Achievement Test (math + reading), constructed as in Krueger (1999).

- Estimate both raw correlations and with controls:
  - quartic in parental household income interacted with marital status
  - parent age at child’s birth
  - parent’s IRA + 401K contributions
  - filing parent’s gender, child’s gender, free lunch status, race
Figure 1: Wage Earnings in 2007 vs. KG Test Score
### Table 2: Test Scores and Earnings in the Cross-Section

<table>
<thead>
<tr>
<th>Dependent Var.:</th>
<th>Wage Earnings</th>
<th>Log Wage Earnings</th>
<th>Wage Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>KG Test percentile</td>
<td>$119.01</td>
<td>$128.44</td>
<td>$81.21</td>
</tr>
<tr>
<td></td>
<td>($6.75)</td>
<td>($8.26)</td>
<td>($8.42)</td>
</tr>
<tr>
<td>KG Test z score</td>
<td>0.174</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental Income Percentile</td>
<td></td>
<td>$119.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>($6.79)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-$326.70</td>
<td>0.070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>($779.30)</td>
<td>(0.086)</td>
<td></td>
</tr>
<tr>
<td>Parental + Demog. Controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Class Fixed Effects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.05</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>Observations</td>
<td>5,609</td>
<td>5,609</td>
<td>5,609</td>
</tr>
</tbody>
</table>

Note: Parental controls are a quartic in parental household income interacted with marital status, parent age at child’s birth, filing parent’s gender, child’s gender, and free lunch
Figure 2a: Percentage Attending College in 2000 vs KG Test Score

The graph shows the relationship between KG Test Score Percentile and College Attendance Rate. There is a positive correlation, indicating that higher KG Test Score Percentiles are associated with higher college attendance rates. The data points are scattered along the trend line, suggesting variability within this trend.
Figure 2a.1: Percentage Ever Attending College vs KG Test Score

- KG Test Score Percentile
- College Attendance Rate

![Graph showing the relationship between KG Test Score Percentile and College Attendance Rate.](image-url)
An Earnings-Based Index of College Quality

- We construct an index of college quality using tax data

- Tuition paid to any higher ed. institution automatically generates a 1098-T form linking student and institution

  - Form filed even if student did not pay out-of-pocket and received a full scholarship

- Find everyone age 20 enrolled in college in 1999

- Calculate average wage earnings in 2007 (from W-2s) by college

- For those who do not attend college, define college quality index as mean earnings for those not in college in 1999
College Mean Wage Earnings by US News Ranking

Mean Earnings at Age 28 vs. US News Rank of College

- $40,000
- $50,000
- $60,000
- $70,000
- $80,000

Graph shows a downward trend with lower US News ranks correlating with higher mean earnings at age 28.
Figure 2b: College Quality in 2000 vs. KG Test Score

The graph shows a positive correlation between KG Test Score Percentile and College Quality (Average Earnings). The earnings increase with higher test score percentiles. The data points are plotted on a linear scale with earnings ranging from $15K to $25K and test score percentiles ranging from 0% to 100%.
Figure 3b: Percentage With 401k by Age 28 vs. KG Test Score

The graph shows the relationship between KG Test Score Percentile and the percentage of individuals with a 401k by age 28. There is a positive correlation, indicating that higher KG Test Score Percentiles are associated with a higher percentage of individuals having a 401k by age 28.
Figure 3c: Average House Value in 2008 Zip Code vs. KG Test Score
Figure 3d: Percentage Married by Age 28 vs. KG Test Score

Percentage Ever Married by Age 28 vs. KG Test Score Percentile

- % Married: 35%, 40%, 45%, 50%, 55%, 60%
- KG Test Score Percentile: 0, 20, 40, 60, 80, 100
Figure 3e: Spouse Earnings vs. KG Test Score
Figure 3f: Summary Outcome Index vs. KG Test Score

![Graph showing the relationship between Summary Outcome Index and KG Test Score Percentile. The graph includes a line of best fit and data points.](image-url)
Part 2: Validity of the STAR Experimental Design

- Threat #1: *Failure of Randomization*
  
  - Prior studies had few baseline measures, limiting ability to evaluate randomization protocol (Schanzenbach 2006)

- We test for balance across class types with an expanded set of parent/sibling characteristics in two ways:
  
  1. Do characteristics vary across small vs. large class types?
  2. Do characteristics vary across classrooms within schools?
<table>
<thead>
<tr>
<th>Dependent Variable: First Obs. Test Score</th>
<th>Small Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Income per Claiming Parent ($1,000s)</td>
<td>0.10 (0.01) [12.52]</td>
</tr>
<tr>
<td>Parent's Age at STAR Birth</td>
<td>0.19 (0.03) [5.57]</td>
</tr>
<tr>
<td>Parent's 401k+IRA Savings ($1,000s)</td>
<td>0.86 (0.19) [4.51]</td>
</tr>
<tr>
<td>Female</td>
<td>3.52 (0.46) [7.71]</td>
</tr>
<tr>
<td>Black</td>
<td>-10.77 (0.90) [-11.98]</td>
</tr>
<tr>
<td>p-Value on F-Statistic</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>10,306</td>
</tr>
</tbody>
</table>

Note: Regressions include school fixed effects.
Validity of the STAR Experimental Design

• Threat #2: Selective Attrition

  • 50% of children starting in KG are missing scores in grade 3; 28% missing scores in grade 8

  • Much less attrition here because we follow 93% of the sample

  • Test for selective attrition in our data in three ways:

    1. Does match rate vary across treatment groups?

    2. Does death rate vary across treatment groups (Muennig et al. 2010)?

    3. Do characteristics of attriters vary across treatment groups?
**Table 5: Match Rates and Death Rates**

<table>
<thead>
<tr>
<th>Dependent Var.:</th>
<th>Matched</th>
<th></th>
<th>Dead</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Small Class Mean</td>
<td>93.68%</td>
<td>93.73%</td>
<td>1.69%</td>
<td>1.69%</td>
</tr>
<tr>
<td>Large Class Mean</td>
<td>94.61%</td>
<td>94.59%</td>
<td>1.59%</td>
<td>1.59%</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.93%</td>
<td>-0.86%</td>
<td>0.10%</td>
<td>0.10%</td>
</tr>
<tr>
<td></td>
<td>(0.55%)</td>
<td>(0.51%)</td>
<td>(0.27%)</td>
<td>(0.27%)</td>
</tr>
<tr>
<td>Class Fixed Effects</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

F-Test of Match Rate on Class Fixed Effects: p value = 0.95
Part 3: Class Size Impacts

- Replicate specifications in previous studies to estimate effect of class size on long-term outcomes

- Independent variable: dummy for small class assignment (ITT)

- Focus on four outcomes:
  1. College attendance in 2000
  2. College quality index
  3. Mean earnings (ages 25-27)
  4. Standardized (SD = 1) summary index of other outcomes:

\[
\text{Index} = 401K + \text{Home Owner} + \text{Married} + \text{Spousal Earnings} + \text{Moved (Leave TN)} + \text{Average Home Value in Zip}
\]
Figure 4b: College Earnings Quality by Class Size

Quintile of College Earnings Index

Percent in Quintile

Lowest
Middle
Highest

Large Class
Small Class
Figure 4c: Effect of Class Size on Wage Earnings by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Large Class</th>
<th>Small Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>$6K</td>
<td>$6K</td>
</tr>
<tr>
<td>2002</td>
<td>$8K</td>
<td>$8K</td>
</tr>
<tr>
<td>2004</td>
<td>$10K</td>
<td>$10K</td>
</tr>
<tr>
<td>2006</td>
<td>$12K</td>
<td>$14K</td>
</tr>
</tbody>
</table>
Table 6: Impacts of Class Size of Adult Outcomes

<table>
<thead>
<tr>
<th>Dependent Var.:</th>
<th>Attended College In 2000</th>
<th>College Mean Earnings</th>
<th>Wage Income</th>
<th>Index of Other Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Class</td>
<td>1.91%</td>
<td>3.38%</td>
<td>$327.5</td>
<td>$79.62</td>
</tr>
<tr>
<td></td>
<td>(0.88%)</td>
<td>(1.48%)</td>
<td>($197.6)</td>
<td>($289.7)</td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Blacks Only</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>10,919</td>
<td>3,914</td>
<td>10,919</td>
<td>10,919</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>26.5%</td>
<td>20.7%</td>
<td>$16,098</td>
<td>$14,792</td>
</tr>
</tbody>
</table>

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, parent age at child’s birth, parent gender, student gender, free-lunch status, and race. Summary index includes a dummy for owning a home, having a 401(k), having moved outside TN, average house value in 2008 zip, a dummy for being married, and spousal earnings (imputing 0 when single). Index is scaled such that a 1 pp. increase in test scores increases the index by 1 unit.
Part 4: Teacher/Peer Effects

• Students randomly assigned to classes that differ in teacher and peer quality

• Do teachers/peers affect adult outcomes?

• Approach 1: Examine impacts of observable characteristics of teachers and peers
  • Begin by examining teacher experience, which predicts test scores

• Throughout remainder of talk, isolate variation across classes within schools and class type (small vs. large)
  • Control for school fixed effects and indicator for class size
Figure 5a: Causal Effect of Teacher Experience on Earnings
Figure 5b: Effects of Teacher Experience on Earnings by Year

Teacher Experience > 9 years

Teacher Experience < 9 years
Table 7: Observable Teacher vs. Peer Effects

<table>
<thead>
<tr>
<th>Dependent Var.:</th>
<th>Wage Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Above Median Teacher Experience</td>
<td>$818.0</td>
</tr>
<tr>
<td>Teacher Experience (years)</td>
<td>$61.93</td>
</tr>
<tr>
<td>Teacher Degree Higher than a BA</td>
<td>$37.47</td>
</tr>
<tr>
<td>% Black Peers</td>
<td>$3,335</td>
</tr>
<tr>
<td>% Female Peers</td>
<td>-$838.6</td>
</tr>
<tr>
<td>% Free-Lunch Peers</td>
<td>-$2,094</td>
</tr>
<tr>
<td>Predicted Peer Scores</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>5,993</td>
</tr>
</tbody>
</table>

Note: All specifications control for school fixed effects and class size, as well demographics and family background.
Class Effects

- We find significant impacts of teacher experience on wages

- Effects of observed peer characteristics are imprecisely estimated

- But most of teacher and peer quality is not captured by these observable measures

  - Similar problem arises in all studies of teacher effects (e.g. Rockoff 2004, Kane and Staiger 2008)

  - To capture these unobservable aspects of class quality, we look directly for class-level effects on wages

  - Class effect includes effect of teachers, peers, and any class-level shocks
Class Effects: ANOVA

• Begin by testing for class effects using analysis of variance

• Do earnings vary across classes by more than what would be predicted by random variation in student abilities?
  
  • F test for significance of class fixed effects
  
  • Random effects estimate of class-level SD for outcomes
Table 8: F-Tests for Kindergarten Class Effects

<table>
<thead>
<tr>
<th></th>
<th>Grade K Scores</th>
<th>Grade 8 Scores</th>
<th>Mean Wage</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value of F-Test on KG Class Fixed Effects</td>
<td>0.000</td>
<td>0.650</td>
<td>0.036</td>
</tr>
<tr>
<td>SD of Class Effects (RE estimate)</td>
<td>8.765</td>
<td>0.000</td>
<td>$1,373</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>School Fixed Effects</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Indicator for Small Class</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Large Classes Only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>5,869</td>
<td>4,470</td>
<td>6,014</td>
</tr>
</tbody>
</table>

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, parent’s age at child’s birth, filing parent’s gender, parent savings, student gender, free-lunch status, and race.
Class Effects on Scores and Earnings

- Key question: Are class effects on KG scores correlated with class effects on earnings?
  - Are improvements in test scores a good proxy for impacts of classes on adult outcomes?
- Cannot simply regress score class f.e.’s on earnings class f.e.’s
  - Class fixed effect includes a student’s own test score → bias toward OLS regression of earnings on scores
  - Equivalent to a weak-instruments problem because class size is finite
  - We address this using a leave-out mean (i.e. jackknife)
Jackknife Estimate of Class Effects

- Define leave-out mean measure of “class quality” as follows:
  - “How good are your classmates’ scores, compared with the classmates you could have had?”

- Class Quality =
  \[
  \text{Mean Peer Scores in Class} - \text{Mean Peer Scores in School}
  \]

- Note that because we are measuring peers’ test scores at the end of KG, class quality reflects teacher + peer effects
- Because students were randomly assigned to classes, class quality varies randomly within schools
Figure 6a: Causal Effect of Class Quality on Own Score
Jackknife Estimate of Class Effects: Placebo Test

- To confirm that jackknife estimate reflects causal effect of class quality on own score, run a placebo test:
  - Predict each student's score from regression of actual score on pre-determined demographics and parental background
  - Test whether class quality is correlated with predicted test score
  - Effectively a test for balance in student characteristics across classes of different quality
Figure 6b: Actual Test Score vs. Predicted Test Score
Figure 6c: Placebo Test: Class Quality and Predicted Own Score
Figure 6d: The Causal Effect of Kindergarten Classrooms on Earnings

Mean Earnings vs. Class Quality (Percentiles)

- Mean Earnings range from -$1,500 to $1,500.
- Class Quality percentiles range from -20 to 20.

The graph shows a positive correlation between class quality and mean earnings, with earnings increasing as class quality increases.
Figure 6e: Placebo Test: Class Quality and Predicted Earnings

Predicted Earnings at Age 28

Class Quality (Percentiles)
Figure 6f: Effect of Class Quality by Year

The graph illustrates the wage earnings trend for Above-Average and Below-Average class quality by year from 2000 to 2006.

- **Wage Earnings**:
  - $6K
  - $8K
  - $10K
  - $12K
  - $14K
  - $16K

- **Year**:
  - 2000
  - 2002
  - 2004
  - 2006

- **Above-Average Class** (Red Line)
- **Below-Average Class** (Blue Line)

Notice the increase in wage earnings over the years for both classes, with Above-Average Class showing a higher trend compared to Below-Average Class.
### Table 9: Class Effects on Test Scores and Earnings

<table>
<thead>
<tr>
<th>Dependent Var.: Test Percentile</th>
<th>Wage Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Test Score</td>
<td>$80.61</td>
</tr>
<tr>
<td>Class Quality</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
</tr>
<tr>
<td>Estimation Method</td>
<td>First Stage</td>
</tr>
<tr>
<td></td>
<td>Form</td>
</tr>
<tr>
<td>Observations</td>
<td>10,306</td>
</tr>
</tbody>
</table>

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, parent’s age at child’s birth, parent’s gender, parent savings, student gender, free-lunch status, and race.
Figure 7: Fadeout of Class Effects: Effect of 1 SD of Class Quality on Test Scores by Grade
Figure 7: Fadeout of Class Effects: Effect of 1 SD of Class Quality on Earnings
Table 10: Class Effects: Impacts on Adult Outcomes

<table>
<thead>
<tr>
<th>Dependent Var.: College in 2000</th>
<th>Attended College in 2000 (Mean Earnings)</th>
<th>College Quality (Mean Earnings)</th>
<th>Index of Other Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Obs. Test</td>
<td>0.20%</td>
<td>$38.33</td>
<td>0.27%</td>
</tr>
<tr>
<td>Score</td>
<td>(0.07%)</td>
<td>($15.38)</td>
<td>(0.16%)</td>
</tr>
<tr>
<td>Individual Controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Observations</td>
<td>10,306</td>
<td>10,306</td>
<td>10,299</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>26.50%</td>
<td>$16,098</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: All specifications control for school fixed effects and class size. Demographic controls include a quartic in parental income interacted with marital status, parent’s age at child birth, parent’s gender, student gender, free-lunch status, and race.
Conclusions

1. Early childhood class effects fade out in test scores but reappear in adult outcomes

2. Contemporaneous test scores are a valid proxy for the benefits of early childhood interventions
   - 1 SD higher in test scores increases earnings by 14.8%
   - Intervention-based estimates similar to OLS with controls

3. Using standard assumptions (Krueger 1999), we estimate that:
   - Small classes $\rightarrow$ $11.5K$ gain in NPV earnings / student
   - +1 SD teacher quality worth $12K / student = $240K / class
   - 1 year of teacher experience worth $34K / class
Cost-Benefit Analysis: Class Size

- Small classes raise test scores by 0.23 SD = $431
  - Assume: 3% real discount rate, constant percent income gains, income follows average US income profile
  - NPV earnings gains of $11,842

- STAR intervention lowered class size by one-third
  - Average school costs = $8,848 per student per year
  - Average treatment = 2.14 years
  - Total Cost per Student = $9,355

- Costs of class size reduction are of the same order of magnitude as earnings gains
Cost-Benefit Analysis: Teacher Quality

• One SD increase in teacher quality raises test scores by 0.2 SD (Rockoff 2004, Kane and Staiger 2008) = $437

  • Assume: 3% real discount rate, constant percent income gains, income follows average US income profile

  → NPV earnings gains of $11,998 per student

• With 20 students per class, would pay $240K extra per year for a 1 SD better teacher

• One year of teacher experience raises earnings by $62

  → NPV gains of $34K per class of 20 students