Appendix A: Algorithm for Matching STAR Records to Tax Data

The tax data were accessed through contract TIRNO-09-R-00007 with the Statistics of Income (SOI) Division at the US Internal Revenue Service. Requests for research contracts by SOI are posted online at the Federal Business Opportunities https://www.fbo.gov/. SOI also welcomes research partnerships between outside academics and internal researchers at SOI.

STAR records were matched to tax data using social security number (SSN), date of birth, gender, name, and STAR elementary school ZIP code. Note that STAR records do not contain all the same information. Almost every STAR record contains date of birth, gender, and last name. Some records contain no SSN while others contain multiple possible SSNs. Some records contain no first name. A missing field yielded a non-match unless otherwise specified.

We first discuss the general logic of the match algorithm and then document the routines in detail. The match algorithm was designed to match as many records as possible using variables that are *not* contingent on ex post outcomes. SSN, date of birth, gender, and last name in the tax data are populated by the Social Security Administration using information that is not contingent on ex post outcomes. First name and ZIP code in tax data are contingent on observing some ex post outcome. First name data derive from information returns, which are typically generated after an adult outcome like employment (W-2 forms), college attendance (1098-T forms), and mortgage interest payment (1098 forms). The ZIP code on the claiming parent's 1040 return is typically from 1996 and is thus contingent on the ex post outcome of the STAR subject not having moved far from her elementary school by age 16.

89.8% of STAR records were matched using only ex ante information. The algorithm first matched as many records as possible using only SSN, date of birth, gender, and last name. It then used first name only to *exclude* candidate matches based on date of birth, gender, and last name, often leaving only one candidate record remaining. Because that exclusion did not condition on an information return having been filed on behalf of that remaining candidate, these matches also did not condition on ex post outcomes.

The match algorithm proceeded as follows, generating seven match types denoted A through G. The matches generated purely through ex-ante information are denoted A through E below and account for 89.8% of STAR records. Matches based on ex-post-information are denoted F and G below and constitute an additional 5.4% of STAR records. The paper reports results using the full 95.0% matched sample, but all the qualitative results hold in the 89.8% sample matched using only ex ante information.

1. Match STAR records to tax records by SSN. For STAR records with multiple possible SSNs, match on all of these SSNs to obtain a set of candidate tax record matches for each STAR record with SSN information. Each candidate tax record contains date of birth, gender, and first four letters of every last name ever assigned to the SSN.

    - Match Type A. Keep unique matches after matching on first four letters of last name, date of birth, and gender.
    - Match Type B. Refine non-unique matches by matching on either first four letters of last name or on "fuzzy" date of birth. Then keep unique matches. Fuzzy date of birth requires the absolute value of the difference between STAR record and tax record dates of birth to be in the set {0,1,2,3,4,5,9,10,18,27} in days, in the set {1,2} in months, or in the set {1} in years. These sets were chosen to reflect common mistakes in recorded dates of birth, such as being off by one day (e.g. 12 vs. 13) or inversion of digits (e.g. 12 vs. 21).

2. Match residual unmatched STAR records to tax records by first four letters of last name, date of birth, and gender.

   - Match Type C. Keep unique matches.
   - Match Type D. Refine non-unique matches by excluding candidates who have a first name issued on information returns (e.g. W-2 forms, 1098-T forms, and various 1099 forms) that does not match the STAR first name on first four letters when the STAR first name is available. Then keep unique matches.
   - Match Type E. Refine residual non-unique matches by excluding candidates who have SSNs that, based on SSN area number, were issued from outside the STAR region (Tennessee and neighboring environs). Then keep unique matches.
   - Match Type F. Refine residual non-unique matches by keeping unique matches after each of the following additional criteria is applied: require a first name match when STAR first name is available, require the candidate tax record's SSN to have been issued from the STAR region, and require the first three digits of the STAR elementary school ZIP code to match the first three digits of the ZIP code on the earliest 1040 return on which the candidate tax record was claimed as a dependent.

3. Match residual unmatched STAR records to tax records by first four letters of last name and fuzzy date of birth.

   - Match Type G. Keep unique matches after each of several criteria is sequentially applied. These criteria include matches on first name, last name, and middle initial using the candidate tax record's information returns; on STAR region using the candidate tax record's SSN area number; and between STAR elementary school ZIP code and ZIP code on the earliest 1040 return on which the candidate tax record was claimed as a dependent.

The seven match types cumulatively yielded a 95.0% match rate:

Match type	Frequency	Percent	Cumulative percent
A	7036	60.8%	60.8%
B	271	2.3%	63.1%
C	699	6.0%	69.2%
D	1391	12.0%	81.2%
E	992	8.6%	89.8%
F	299	2.6%	92.4%
G	304	2.6%	95.0%

Identifiers such as names and SSN's were used solely for the matching procedure. After the match was completed, the data were de-identified (i.e., individual identifiers such as names and SSNs were stripped) and the statistical analysis was conducted using the de-identified dataset.

Appendix B: Derivations for Measurement of Unobserved Class Quality

This appendix derives the estimators discussed in the empirical model in Section V and quantifies the degree of attenuation and reflection bias. We first use equations (**??**) and (**??**) to define average of test scores and earnings within each class c and school n:

$$\begin{aligned}
s_{cn} &= d_n + z_{cn} + a_{cn} \\
y_{cn} &= \delta_n + \beta z_{cn} + z_{cn}^Y + \rho a_{cn} + \nu_{cn} \\
s_n &= d_n + z_n + a_n \\
y_n &= \delta_n + \beta z_n + z_n^Y + \rho a_n + \nu_n.
\end{aligned}$$

We define variables demeaned within schools as

$$\begin{aligned}
s_{icn} - s_n &= z_{cn} - z_n + a_{icn} - a_n \\
\Delta s_{cn} \equiv s_{cn} - s_n &= z_{cn} - z_n + a_{cn} - a_n, \\
y_{icn} - y_n &= \beta(z_{cn} - z_n) + (z_{icn}^Y - z_n^Y) + \rho(a_{icn} - a_n) + \nu_{icn} - \nu_n \\
y_{cn} - y_n &= \beta(z_{cn} - z_n) + (z_{cn}^Y - z_n^Y) + \rho(a_{cn} - a_n) + \nu_{cn} - \nu_n.
\end{aligned}$$

Recall that a_{icn} and ν_{icn} are independent of each other and z_{cn}. Let $\sigma^2 = var(a_{icn})$. We assume in parts 1 and 2 below that $z_{cn}, z_{cn}^Y \perp a_{icn}$, ruling out peer effects. Note also that, as $z_{cn} \perp z_{cn}^Y$, the component of classroom environments that affects only test scores drops out entirely of the covariance analysis below. In what follows, we take the number of students per class I and the number of classrooms per school C as fixed and analyze the asymptotic properties of various estimators as the number of schools $N \to \infty$.

1. Mean score estimator. The simplest proxy for class quality is the average test score within a class. Since we include school fixed effects in all specifications, s_{cn} is equivalent to Δs_{cn} as defined above. Therefore, consider the following (school) fixed effects OLS regression:

$$y_{icn} = \alpha_n + b^M \Delta s_{cn} + \varepsilon_{icn}. \tag{1}$$

As the number of schools $N \to \infty$, the coefficient estimate $\hat{b}^M$ converges to

$$plim_{\ N \longrightarrow \infty}\ \hat{b}^M = \frac{cov(y_{icn} - y_n, s_{cn} - s_n)}{var(s_{cn} - s_n)},$$

which we can rewrite as

$$\begin{aligned}
plim_{\ N \longrightarrow \infty}\ \hat{b}^M &= \frac{cov\left(\beta(z_{cn} - z_n) + \rho(a_{icn} - \frac{\sum_k \sum_j a_{jkn}}{I \cdot C}), z_{cn} - z_n + \frac{\sum_j a_{jcn}}{I} - \frac{\sum_k \sum_j a_{jkn}}{I \cdot C}\right)}{var\left(z_{cn} - z_n + \frac{\sum_j a_{jcn}}{I} - \frac{\sum_k \sum_j a_{jkn}}{I \cdot C}\right)} \\
&= \frac{\beta var(z_{cn} - z_n) + \rho \sigma^2 \frac{C-1}{IC}}{var(z_{cn} - z_n) + \sigma^2 \frac{C-1}{IC}}.
\end{aligned}$$

Even absent class effects ($\beta = 0$), we obtain $plim_{\ N \longrightarrow \infty}\ \hat{b}^M > 0$ if I is finite and $\rho > 0$. With finite class size, b^M is upward-biased due to the correlation between wages and own-score, which is included within the class quality measure.

2. Leave-out mean estimator. We address the upward bias due to the own observation problem using a leave-out mean estimator. Consider the OLS regression with school fixed effects

$$y_{icn} = \alpha_n + b^{LM} \Delta s_{cn}^{-i} + \varepsilon_{icn}. \tag{2}$$

where $\Delta s_{cn}^{-i} = s_{cn}^{-i} - s_n^{-i}$ is defined as in equation (??). The coefficient b^{LM} converges to

$$plim_{\ N \longrightarrow \infty}\ \hat{b}^{LM} = \frac{cov(y_{icn} - y_n, s_{cn}^{-i} - s_n^{-i})}{var(s_{cn}^{-i} - s_n^{-i})},$$

which we can rewrite as

$$\begin{aligned} plim_{\ N \longrightarrow \infty}\ \hat{b}^{LM} &= \frac{cov\left(\beta(z_{cn} - z_n) + \rho(a_{icn} - a_n), \frac{IC}{IC-1}(z_{cn} - z_n) + \frac{1}{I-1}\sum_{j \neq i} a_{jcn} - \frac{1}{IC-1}\sum_k \sum_{j \neq i} a_{jkn}\right)}{var\left(\frac{IC}{IC-1}(z_{cn} - z_n) + \frac{1}{I-1}\sum_{j \neq i} a_{jcn} - \frac{1}{IC-1}\sum_k \sum_{j \neq i} a_{jkn}\right)} \\ &= \beta \times \frac{\frac{IC}{IC-1} var(z_{cn} - z_n)}{\frac{(IC)^2}{(IC-1)^2} var(z_{cn} - z_n) + \frac{\sigma^2}{I-1} - \frac{\sigma^2}{I \cdot C-1}} \end{aligned}$$

Hence, $plim_{\ N \longrightarrow \infty}\ \hat{b}^{LM} = 0$ if and only if $\beta var(z_{cn} - z_n) = 0$ (no class effects) even when I and C are finite.* However, b^{LM} is attenuated relative to β because peer scores are a noisy measure of class quality.

Quantifying the degree of attenuation bias. We can quantify the degree of attenuation bias by using the within-class variance of test scores as an estimate of $\sigma^2 = var(a_{icn})$. First, note that:

$$\begin{aligned} \widehat{var}(z_{cn} - z_n) &= \frac{(IC-1)^2}{(IC)^2}\left[\widehat{var}\left(s_{cn}^{-i} - s_n^{-i}\right) - \left(\frac{\hat{\sigma}^2}{I-1} - \frac{\hat{\sigma}^2}{I \cdot C - 1}\right)\right] \\ &= \frac{(83.63)^2}{(84.73)^2}\left[81.75 - \left(\frac{437.4}{19.07} - \frac{437.4}{83.63}\right)\right] \\ &= 62.39 \end{aligned}$$

where we use the sample harmonic means for IC, $IC-1$, and $I-1$ because the number of students in each class and school varies across the sample. This implies an estimate of bias of

$$\frac{\frac{83.63}{84.73} 62.39}{\frac{(83.63)^2}{(84.73)^2} 62.39 + \frac{437.4}{19.07} - \frac{437.4}{83.63}} = 0.773.$$

That is, b^{LM} is attenuated relative to β by 22.7%. Note that this bias calculation assumes that all students in the class were randomly assigned, which is true only in KG. In later grades, the degree of attenuation in b^{LM} when equation (2) is estimated using new entrants is larger than 22.7%, because existing students were not necessarily re-randomized at the start of subsequent grades.

3. Peer effects and reflection bias. With peer effects, the assumption $z_{cn} \perp a_{icn}$ does not hold. We expect z_{cn} and a_{icn} to be positively correlated with peer effects as a higher ability student has a positive impact on the class. This leads to an upward bias in both b^{LM} and b^{SS} due to the reflection problem. To characterize the magnitude of this bias, consider a standard

*The leave-out mean estimator b^{LM} is consistent as the number of schools grows large, but is downward biased in small samples because own scores s_{icn} and peer scores Δs_{cn}^{-i} are mechanically negatively correlated within classrooms. Monte Carlo simulations suggest that this finite sample bias is negligible in practice with the number of schools and classrooms in the STAR data.

linear-in-the-means model of peer effects, in which

$$z_{cn} = t_{cn} + \frac{\theta}{I}\sum_j a_{jcn}$$

with $t_{cn} \perp a_{jcn}$ for all j. Here t_{cn} represents the component of class effects independent of peer effects (e.g., a pure teacher effect). The parameter $\theta > 0$ captures the strength of peer effects. Averaging across classrooms within a school implies that

$$z_n = t_n + \frac{\theta}{IC}\sum_k \sum_j a_{jkn}.$$

In this model, the leave-out mean proxy of class quality is

$$\Delta s_{cn}^{-i} = s_{cn}^{-i} - s_n^i = \frac{IC}{IC-1}(t_{cn} - t_n) + \theta\frac{IC}{IC-1}(a_{cn} - a_n) + \frac{1}{I-1}\sum_{j\neq i} a_{jcn} - \frac{1}{IC-1}\sum_k \sum_{j\neq i} a_{jkn}$$

and as N grows large $\hat{b}^{LM}$ converges to

$$\begin{aligned} plim_{\ N\longrightarrow\infty}\ \hat{b}^{LM} &= \frac{cov(y_{icn} - y_n, s_{cn}^{-i} - s_n^{-i})}{var(s_{cn}^{-i} - s_n^{-i})} \\ &= \frac{\beta\cdot\left[\frac{IC}{IC-1}var(t_{cn} - t_n) + (\theta+\theta^2)\sigma^2\frac{C-1}{IC-1}\right] + \rho\theta\sigma^2\frac{C-1}{IC-1}}{\frac{(IC)^2}{(IC-1)^2}var(t_{cn} - t_n) + (2\theta+\theta^2)\sigma^2\frac{IC(C-1)}{(IC-1)^2} + \frac{\sigma^2}{I-1} - \frac{\sigma^2}{I\cdot C-1}} \end{aligned}$$

The last term in the numerator is the reflection bias that arises because a high ability student has both high earnings (through ρ) and a positive impact on peers' scores (through θ). Because of this term, we can again obtain $plim_{\ N\longrightarrow\infty}\ \hat{b}^{LM} > 0$ even when $\beta = 0$. This bias occurs iff $\theta > 0$ (i.e., we estimate $b^{LM} > 0$ only if there are peer effects on test scores). This bias is of order $\frac{1}{I}$ since any given student is only one of I students in a class that affects class quality.

Bounding the degree of reflection bias. We use the estimated impact of KG class quality on 8th grade test scores to bound the degree of reflection bias in our estimate of the impact of class quality on earnings. Recall that the reflection bias arises because a high ability student has better long-term outcomes and also has a positive impact on peers' kindergarten test scores. Therefore, the same reflection bias is present when estimating $\hat{b}^{LM}$ using eighth grade test scores as the outcome instead of earnings.

Denote by $\hat{b}_e^{LM}$ the estimated coefficient on Δs_{cn}^{-i} when the outcome y is earnings and $\hat{b}_s^{LM}$ the same coefficient when the outcome y is grade 8 test scores.[†] Similarly, denote by ρ_e and ρ_s the (within class) correlation between individual kindergarten test score and earnings or eighth grade test score. Under our parametric assumptions, these two parameters can be estimated by an OLS regression $y_{icn} = \alpha_{cn} + \rho s_{icn} + \varepsilon_{icn}$ that includes class fixed effects.

To obtain an upper bound on the degree of reflection bias, we make the extreme assumption that the effect of kindergarten class quality on eighth grade test scores ($\hat{b}_s^{LM}$) is due entirely to the reflection bias. If there are no pure class effects ($var(t_{cn} - t_n) = 0$) and peers do not affect earnings

[†]The latest test score we have in our data is in grade 8. We find similar results if we use other grades, such as fourth grade test scores.

($\beta = 0$),

$$\text{(3)} \qquad plim\, \hat{b}^{LM} = \frac{\rho\theta}{\frac{1}{1-\frac{1}{I}} + \frac{2\theta+\theta^2}{1-\frac{1}{IC}}} \simeq \frac{\rho\theta}{(1+\theta)^2}$$

Using equation (3) for $\hat{b}_s^{LM}$ and the estimate of $\hat{\rho}_s$, we obtain an estimate of the reflection bias parameter $\frac{\theta}{(1+\theta)^2} = \hat{b}_s^{LM}/\hat{\rho}_s$. Combining this estimate and the estimate $\hat{\rho}_e$, we can then use equation (3) for $\hat{b}_e^{LM}$ to obtain an upper bound on the $\hat{b}_e^{LM}$ that could arise solely from reflection bias.

We implement the bound empirically by estimating the relevant parameters conditional on the vector of parent and student demographics, using regression specifications that parallel those used in column 3 of Table IV and column 2 of Table VIII. For eighth grade scores, we estimate $\hat{b}_s^{LM} = 0.057$ (SE = 0.036) and $\rho_s = 0.597$ (SE = 0.016), and hence

$$\frac{\theta}{(1+\theta)^2} = \frac{0.057}{0.597} = 0.0955.$$

For earnings, we estimate $\rho_e = \$90.04$ (SE = \$8.65) in Table IV. Hence, if the entire effect of class quality on earnings were due to reflection bias, we would obtain

$$\hat{b}_e^{LM} = \frac{\rho_e\theta}{(1+\theta)^2} = \$90.04 \cdot 0955 = \$8.60 \text{ (SE} = \$5.49)$$

where the standard error is computed using the delta method under the assumption that the estimates of $\hat{b}_s^{LM}$, ρ_s, and ρ_e are uncorrelated. This upper bound of \$8.60 due to reflection bias is only 17% of the estimate of $\hat{b}_e^{LM} = \$50.61$ (SE = \$17.45) in Table VIII. Note that the degree of reflection bias would be smaller in the presence of class quality effects ($\beta > 0$); hence, 17% is an upper bound on the degree of reflection bias in a linear-in-means model of peer effects.

Appendix C: Cost-Benefit Analysis

We make the following assumptions to calculate the benefits of the policies considered in the conclusion. First, following Krueger (1999), we assume a 3% annual discount rate and discount all earnings streams back to age 6, the point of the intervention. Second, we use the mean wage earnings of a random sample of the U.S. population in 2007 as a baseline earnings profile over the lifecycle. Third, because we can observe earnings impacts only up to age 27, we must make an assumption about the impacts after that point. We assume that the percentage gain observed at age 27 remains constant over the lifecycle. This assumption may understate the total benefits because the earnings impacts appear to grow over time, for example as college graduates have steeper earnings profiles. Finally, our calculations ignore non-monetary returns to education such as reduced crime. They also ignore general equilibrium effects: increasing the education of the population at large would increase the supply of skilled labor and may depress wage rates for more educated individuals, reducing total social benefits. Under these assumptions, we calculate the present-value earnings gains for a classroom of 20 students from three interventions: improvements in classroom quality, reductions in class size, and improvements in teacher quality.

(1) Class Quality. The random-effects estimate reported in column 4 of Table VII implies that increasing class quality by one standard deviation of the distribution within schools raises earnings by \$1,520 (9.6%) at age 27. Under the preceding assumptions, this translates into a lifetime earnings gain of approximately \$39,100 for the average individual. This implies a present-value benefit of \$782,000 for improving class quality by one within-school standard deviation.

(2) Class Size. We calculate the benefits of reducing class size by 33% in two ways. The first method uses the estimated earnings gain from being assigned to a small class reported in column 5 of Table V. The point estimate of $4 in Table V translates into a lifetime earnings gain from reducing class size by 33% for one year of $103 in present value per student, or $2,057 for a class that originally had twenty students. But this estimate is imprecise: the 95% confidence interval for the lifetime earnings gain of reducing class size by 33% for one year ranges from -$17,500 to $17,700 per child. Moreover, the results for other measures such as college attendance suggest that the earnings impact may be larger in the long run.

To obtain more precise estimates, we predict the gains from class size reduction using the estimated impact of classroom quality on scores and earnings. We estimate that a 1 percentile increase in class quality raises test scores by 0.66 percentiles and earnings by $50.6. This implies an earnings gain of $76.67 per percentile (or 13.1% per standard deviation) increase in test scores. We make the strong assumption that the ratio of earnings gains to test score gains is the same for changes in class size as it is for improvements in class quality more generally.[‡] Under this assumption, smaller classes (which raised test scores by 4.8 percentiles) are predicted to raise earnings by $4.8 \times \$76.7 = \$368$ (2.3%) at age 27. This calculation implies a present value earnings gain from class size reduction of $9,460 per student and $189,000 for the classroom.

Calculations analogous to those in Krueger (1999) imply that the average cost per child of reducing class size by 33% for 2.14 years (the mean treatment duration for STAR students) is $9,355 in 2009 dollars.[§] Our second calculation suggests that the benefit of reducing class size might outweigh the costs. However, we must wait for more time to elapse before we can determine whether the predicted earnings gains based on the class quality estimates are in fact realized by those who attended smaller classes.

(3) Teachers. We calculate the benefits of improving teacher quality in two ways. The first method uses the estimated earnings gain of $57 from being assigned to a kindergarten teacher with one year of extra experience, reported in Figure IIIb. The standard deviation of teacher experience in our sample is 5.8 years. Hence, a one standard deviation increase in teacher experience raises earnings by $331 (2.1%) at age 27. This translates into a lifetime earnings gain of $8,500 in present value, or $170,000 for a class of twenty students.

The limitation of the preceding calculation is that it is based upon only one observable aspect of teacher quality. To incorporate other aspects of teacher quality, we again develop a prediction based on the impacts of class quality on scores and earnings. Rockoff (2004), Rivkin, Hanushek, and Kain (2005), and Kane and Staiger (2008) use datasets with repeated teacher observations to estimate that a one standard deviation increase in teacher quality raises test scores by approximately 0.2 standard deviations (5.4 percentiles). Under the strong assumption that the ratio of earnings gains to test score gains is the same for changes in teacher quality and class quality more broadly, this translates into an earnings gain of $5.4 \times \$76.7 = \$416$ (2.6%) at age 27. This implies a present-value earnings gain of $10,700 per student. A one standard deviation improvement in teacher quality in a single year generates earnings gains of $214,000 for a class of twenty students.

[‡]This assumption clearly does not hold for all types of interventions. As an extreme example, raising test scores by cheating would be unlikely to yield an earnings gain of $77 per percentile improvement in test scores. The $77 per percentile measure should be viewed as a prior estimate of the expected gain when evaluating interventions such as class size or teacher quality for which precise estimates of earnings impacts are not yet available.

[§]This cost is obtained as follows. The annual cost of school for a child is $8,848 per year. Small classes had 15.1 students on average, while large classes had 22.56 students on average. The average small class treatment lasted 2.14 years. Hence, the cost per student of reducing class size is (22.56/15.1-1)*2.14*8848= $9,355.

APPENDIX TABLE I
Correlations of Earnings Over the Life Cycle

Age	Correlation between Wage Earnings at Age x and x+6
18	0.36
19	0.36
20	0.37
21	0.41
22	0.47
23	0.55
24	0.60
25	0.62
26	**0.65**
27	0.67
28	0.69
29	0.70
30	0.71
31	0.72
32	0.74
33	0.75
34	0.75
35	0.77
36	0.77
37	0.78
38	0.79
39	0.79
40	0.80
41	0.80
42	0.81
43	0.81
44	0.81
45	0.81
46	0.80
47	0.80
48	0.80
49	0.79
50	0.78

Notes: This table presents correlations between individual mean wage earnings 1999-2001 and individual mean wage earnings 2005-2007 (including zeros for people with no wage earnings) for different ages in a 3% random sample of the US population. Age is defined as age on December 31, 2000. Individuals with mean wage earnings greater than $200,000 over years 1999-2001 or 2005-2007 are omitted. The earnings outcome most commonly used in the tables is STAR subject mean wage earnings 2005-2007. The typical STAR subject was 26 on December 31, 2006. The row in bold implies that STAR subjects' mean wage earnings 2005-2007 are predicted to correlate with their mean wage earnings 2011-2013 (when STAR subjects are approximately aged 31-33) with a coefficient of 0.65.

APPENDIX TABLE II
Randomization Tests by Entry Grade

	p-value				
Entry Grade:	Grade K (1)	Grade 1 (2)	Grade 2 (3)	Grade 3 (4)	Pooled (5)
Parent's income ($1000s)	0.848	0.081	0.127	0.117	0.412
Mother's age at STAR birth	0.654	0.082	0.874	0.555	0.165
Parents have 401 (k)	0.501	0.427	0.634	0.567	0.957
Parents married	0.820	0.921	0.981	0.280	0.116
Parents own home	0.435	0.075	0.158	0.879	0.929
Student black	0.995	1.000	0.939	0.997	0.040
Student free-lunch	0.350	0.060	0.159	0.798	0.469
Student's age at KG entry	0.567	0.008	0.251	0.972	0.304
Student female	0.502	0.413	0.625	0.069	0.052
Predicted earnings	0.916	0.674	0.035	0.280	0.645

Notes: Each cell in Columns 1-4 reports the p value on an F test for joint significance of classroom fixed effects in a separate OLS regression. The row indicates the dependent variable. The column indicates the sample of entrants used. Each regression includes school fixed effects, so one classroom fixed effect per school is omitted. In Column 5, we pool all entry grades by regressing a student's own characteristic on the difference between his classmates' and grade-specific schoolmates' mean values of that characteristic. Each row of Column 6 reports p values from a separate regression that includes school-by-entry-grade fixed effects. The p values are from a t test for the significance of the coefficient on peer characteristics, i.e. a test for significant intra-class correlation in the variable listed in each row.

APPENDIX TABLE III

Correlation Between Test Scores and Components of Summary Index

Dependent Variable:	Home Owner	Have 401 K	Married	Moved Out of State	College Grads in 2007 Zip
	(%)	(%)	(%)	(%)	(%)
A. All Entrants	(1)	(2)	(3)	(4)	(5)
Entry-grade test percentile	0.159	0.109	0.057	0.179	0.053
	(0.022)	(0.025)	(0.025)	(0.025)	(0.006)
Observations	9,939	9,939	9,939	9,301	9,424
B. KG Entrants					
KG test percentile	0.136	0.100	0.048	0.145	0.053
	(0.028)	(0.035)	(0.035)	(0.021)	(0.007)
Observations	5,621	5,621	5,621	5,354	5,367
C. 1st Grade Entrants					
1st grade test percentile	0.205	0.113	0.076	0.282	0.046
	(0.050)	(0.047)	(0.050)	(0.053)	(0.012)
Observations	2,124	2,124	2,124	1,934	2,002
D. 2nd Grade Entrants					
2nd grade test percentile	0.089	0.072	0.080	0.226	0.059
	(0.072)	(0.080)	(0.082)	(0.105)	(0.024)
Observations	1,215	1,215	1,215	1,112	1,147
E. 3rd Grade Entrants					
3rd grade test percentile	0.231	0.196	0.022	0.070	0.056
	(0.105)	(0.085)	(0.101)	(0.098)	(0.021)
Observations	979	979	979	901	908

Notes: This table replicates the specification of Column 9 of Table IV for various subgroups and the five constituent components of summary index. Each row specifies the sample restriction according to the year that the student entered a STAR school. Each column specifies which of the five components of the summary index is used as the dependent variable. See Table I for definitions of each outcome variable. See notes to Table IV for the regression specification. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE IV
Correlation between Test Scores and Adult Outcomes by Entry Grade

Dependent Variable:	Wage Earnings				College in 2000	College by Age 27	College Quality	Summary Index
	($)	($)	($)	($)	(%)	(%)	(%)	(% SD)
A. KG Entrants	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KG test percentile	131.7	93.79	-8.529	105.5	0.398	0.527	36.21	0.492
	(12.24)	(11.63)	(15.30)	(11.39)	(0.031)	(0.029)	(4.08)	(0.058)
8th grade test percentile			156.4					
			(12.33)					
Parental income percentile				157.7				
				(9.57)				
B. 1st Grade Entrants								
1st grade test percentile	134.8	80.38	12.70	82.34	0.292	0.449	17.61	0.654
	(15.09)	(14.81)	(23.82)	(14.81)	(0.041)	(0.047)	(4.71)	(0.096)
8th grade test percentile			124.8					
			(27.92)					
Parental income percentile				136.7				
				(18.15)				
C. 2nd Grade Entrants								
2nd grade test percentile	150.3	67.29	-42.16	65.70	0.308	0.459	42.02	0.568
	(19.18)	(25.56)	(43.88)	(25.23)	(0.064)	(0.076)	(11.59)	(0.153)
8th grade test percentile			183.7					
			(47.40)					
Parental income percentile				112.9				
				(22.97)				
D. 3rd Grade Entrants								
3rd grade test percentile	146.2	99.03	87.60	102.2	0.347	0.534	28.09	0.589
	(19.80)	(31.34)	(50.63)	(30.00)	(0.070)	(0.088)	(7.48)	(0.183)
8th grade test percentile			58.54					
			(56.81)					
Parental income percentile				99.08				
				(26.84)				
Class fixed effects		x	x	x	x	x	x	x
Demographic controls		x	x		x	x	x	x

Notes: This table replicates the specifications of Columns 1 and 3-9 of Table IV for four subgroups, one for each year of entering students. See notes to that table for definitions of variables and regression specifications. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE V
Correlation Between Test Scores and Earnings by Grade

Dependent Variable:	Wage Earnings ($)	
	(1)	(2)
Grade K scores	93.79	39.46
	(11.63)	(30.08)
Grade 1 scores	98.40	71.21
	(11.83)	(23.53)
Grade 2 scores	88.72	83.32
	(14.15)	(24.66)
Grade 3 scores	97.43	103.5
	(12.78)	(20.08)
Grade 4 scores	94.71	91.40
	(12.40)	(20.69)
Grade 5 scores	110.8	113.2
	(10.81)	(19.24)
Grade 6 scores	121.6	139.5
	(11.37)	(21.53)
Grade 7 scores	138.2	158.7
	(11.53)	(21.82)
Grade 8 scores	148.9	155.2
	(11.11)	(21.47)
Sample	All KG Entrants	Constant Sample of KG Entrants

Notes: Each row of this table reports the coefficient on test score from a separate regression that replicates Column 2 of Table IV, replacing KG test score with the test score from the listed grade. The STAR data do not contain test scores for all students in all grades. Regressions underlying Column 1 use all students who entered a STAR school in kindergarten and who have a test score for the listed grade. Regressions underlying Column 2 use only these KG entrants who have a test score for every grade. The coefficients in Column 1 are used to construct Figure VIb. See notes to Table IV for other variable definitions and the regression specification. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE VI
Correlation between Test Scores and Adult Outcomes: Heterogeneity Analysis

Dependent Variable:	Wage Earnings	College in 2000	College by Age 27	College Quality	Summary Index
	($)	(%)	(%)	($)	(% of SD)
	(1)	(2)	(3)	(4)	(5)
Blacks	105.8	0.300	0.526	27.39	0.553
	(14.06)	(0.044)	(0.036)	(4.89)	(0.091)
Whites	83.06	0.392	0.504	34.21	0.537
	(12.24)	(0.028)	(0.030)	(4.19)	(0.061)
Males	77.16	0.323	0.480	28.17	0.415
	(11.48)	(0.034)	(0.034)	(3.87)	(0.074)
Females	114.6	0.404	0.539	35.72	0.663
	(12.44)	(0.040)	(0.041)	(4.98)	(0.086)
Free-lunch eligible	87.28	0.255	0.429	17.13	0.513
	(9.06)	(0.031)	(0.032)	(2.78)	(0.064)
Not elig. for free lunch	94.70	0.544	0.618	58.02	0.631
	(20.01)	(0.041)	(0.041)	(6.70)	(0.091)

Notes: This table replicates selected specifications of Table IV for various subgroups of students. Each cell reports the coefficient on entry-grade test score percentile from a separate OLS regression limited to the sub-group defined in the row with the dependent variable defined in the column header. Each column 1-5 uses the specification from the following column of Table IV, respectively: 3, 6, 7, 8, and 9. Free-lunch eligible is an indicator for whether the student was ever eligible for free or reduced-price lunch during the experiment. See notes to Table IV for regressions specifications and other variable definitions. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE VII
Effect of Clustering on Standard Errors

Dependent Var:	Grade K Test Score		Earnings ($)		
	Krueger (1999), Table V col 6	Linked STAR-IRS	Linked STAR-IRS	Linked STAR-IRS	Linked STAR-IRS
	(1)	(2)	(3)	(4)	(5)
Small Class Dummy	5.37	5.40	-124		
SE w/o clustering		(0.78)	(325)		
SE cluster by class	(1.25)	(1.29)	(299)		
SE cluster by school		(1.45)	(336)		
High Teacher Exp.				1093	
SE w/o clustering				(453)	
SE cluster by class				(437)	
SE cluster by school				(545)	
Class Quality					50.61
SE w/o clustering					(15.35)
SE cluster by class					(14.76)
SE cluster by school					(17.45)
Demographic controls			x	x	x
Observations	5,861	5,869	10,992	6,005	10,959

Notes: Table shows estimates from regressions with standard errors calculated in three ways: no clustering, clustering by entry classroom, and clustering by school. In Columns 1-2, independent variables include indicators for assignment to small class as well as assignment to regular class with aide (coefficient not shown) to replicate the specification in Krueger (1999). Columns 3-5 replicate the key earnings specifications from Table V (Col. 5), Table VI (Col. 1) and Table VII (Col. 2). Columns 1-2 use grade K entrants, while Columns 3-5 use all matched students regardless of entry grade. All columns include school by entry wave fixed effects.

APPENDIX TABLE VIII

Effects of Class Size on Adult Outcomes: Kindergarten Entrants Only

Dependent Variable:	Test Score	College in 2000	College by Age 27	College Quality	Wage Earnings	Summary Index
	(%)	(%)	(%)	($)	($)	(% of SD)
	(1)	(2)	(3)	(4)	(5)	(6)
Small class (no controls)	5.37	1.56	1.57	165	-3.23	6.07
	(1.26)	(1.50)	(1.29)	(145)	(431)	(2.63)
Small class (with controls)	5.16	1.70	1.64	185	-57.6	5.64
	(1.21)	(1.35)	(1.22)	(143)	(440)	(2.60)
Observations	5,621	6,025	6,025	6,025	6,025	6,025
Mean of dep. var.	51.44	31.47	51.45	27,422	17,111	4.88

Notes: This table replicates Table V, restricting the sample to students who entered a STAR school in kindergarten. See notes to Table V for regression specifications and variable definitions. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE IX
Results for Alternative Measures of Wage Earnings

Dependent Variable:	Positive Mean Earnings	Above Median Earnings	Percentile Earnings	Household Income	2007 Wages
	(%)	(%)	(%)	($)	($)
A. Cross-Sectional Correlations	(1)	(2)	(3)	(4)	(5)
Entry-grade test percentile	0.055	0.222	0.156	126.9	101.0
	(0.021)	(0.028)	(0.017)	(11.75)	(9.735)
B. Class Size Impacts					
Small class	0.123	0.482	-0.191	241.5	-263.3
	(0.741)	(1.213)	(0.613)	(457.2)	(404.7)
C. Class Quality Impacts					
Class quality (peer scores)	0.062	0.176	0.098	52.40	45.14
	(0.032)	(0.052)	(0.031)	(20.19)	(19.05)
Mean of dep. var.	86.14	50.00	50.00	23,883	16,946

Notes: This table replicates certain specifications using alternative measures of earnings outcomes. Panel A replicates the specification of Column 3 of Table IV. Panel B replicates the "with controls" specification of Row 2 of Table V. Panel C replicates the specification of Column 2 of Table VIII. Each of the five columns denotes a different earnings variable used in that column's regressions: (1) an indicator for having positive wage earnings in any year 2005-2007, (2) an indicator for having average wage earnings over years 2005-2007 greater than the sample median ($12,553), (3) the within-sample percentile of a student's average wage earnings over years 2005-2007, (4) total household income for each student over years 2005-2007, defined as adjusted gross income adjusted for tax-exempt Social Security and interest payments, and (5) wage earnings in 2007, winsorized at $100,000. See notes to Tables IV, V, and VIII for regression specifications and other variable definitions. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE X

Impacts of Class Size and Quality on Components of Summary Outcome Index

Dependent Variable:	Home Owner	Have 401 K	Married	Moved Out of State	College Grads in 2007 Zip	Predicted Earnings Summary Index
	(%)	(%)	(%)	(%)	(%)	($)
A. Class Size	(1)	(2)	(3)	(4)	(5)	(6)
Small class	0.712	2.888	1.872	1.105	-0.038	438.9
	(1.039)	(1.042)	(1.082)	(0.930)	(0.223)	(209.8)
Observations	10,992	10,992	10,404	10,268	10,404	10,992
B. Class Quality						
Class quality (peer scores)	0.080	0.053	0.056	0.029	0.019	16.48
	(0.047)	(0.058)	(0.051)	(0.045)	(0.012)	(12.16)
Observations	10,959	10,959	10,375	10,238	10,375	10,959
Mean of dep. var.	30.80	28.18	44.83	27.53	17.60	15,912

Notes: Columns 1-5 decompose the impacts of class size and quality on the summary index into impacts on each of the summary index's five constituent components. Panel A replicates the "with controls" specification of Row 2 of Table V for each component. Panel B replicates Column 9 of Table VIII for each component. See notes to those tables for regression specifications and sample definitions. See Table I for definitions of the dependent variables used in Columns 1-5, each of which is a component of summary index. Column 6 reports impacts on an alternative "predicted earnings" summary index. This index is constructed by predicting earnings from a regression of mean wage earnings over years 2005-2007 on the five dependent variables in Columns 1-5. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE XI

Impacts of Class Size and Quality: Heterogeneity Analysis

Dependent Variable:	Test Score	College in 2000	College by Age 27	College Quality	Wage Earnings	Summary Index
	(%)	(%)	(%)	($)	($)	(% of SD)
A. Effect of Small Class	(1)	(2)	(3)	(4)	(5)	(6)
Blacks	6.871	2.722	5.312	249.0	250.0	6.308
	(1.825)	(2.036)	(2.417)	(134.1)	(540.0)	(3.343)
Whites	3.699	1.065	-0.177	38.94	-348.1	4.388
	(1.109)	(1.099)	(1.120)	(140.3)	(413.6)	(2.649)
Males	4.883	2.594	2.279	244.5	798.3	10.89
	(1.103)	(1.278)	(1.414)	(127.0)	(497.7)	(3.188)
Females	4.360	0.716	0.454	6.638	-1130	-2.599
	(1.226)	(1.611)	(1.621)	(163.3)	(434.8)	(3.069)
Free-lunch eligible	5.767	0.837	3.908	-2.517	-251.7	3.162
	(1.299)	(1.242)	(1.560)	(88.34)	(394.7)	(2.790)
Not elig. for free lunch	3.376	3.592	-0.914	296.6	293.0	7.292
	(1.288)	(1.691)	(1.480)	(222.8)	(595.5)	(3.390)
B. Effect of Class Quality (peer scores)						
Blacks	0.732	0.081	0.089	3.197	36.22	0.236
	(0.027)	(0.066)	(0.086)	(7.735)	(18.60)	(0.099)
Whites	0.582	0.069	0.075	12.16	68.17	0.180
	(0.041)	(0.070)	(0.068)	(6.794)	(29.02)	(0.181)
Males	0.654	0.016	0.067	8.257	66.03	0.357
	(0.033)	(0.061)	(0.068)	(7.548)	(25.48)	(0.158)
Females	0.654	0.151	0.131	7.845	32.37	0.032
	(0.039)	(0.060)	(0.065)	(6.162)	(19.53)	(0.163)
Free-lunch eligible	0.650	0.058	0.089	2.319	53.14	0.077
	(0.036)	(0.052)	(0.068)	(3.906)	(17.55)	(0.120)
Not elig. for free lunch	0.652	0.149	0.104	22.34	47.11	0.442
	(0.043)	(0.103)	(0.093)	(11.02)	(39.64)	(0.213)

Notes: Each cell reports a coefficient estimate from a separate OLS regression. Panel A replicates the "with controls" specification of Row 2 of Table V, for various subgroups and dependent variables. Panel B replicates the specification of Column 2 of Table VIII, for various subgroups and dependent variables. Free-lunch eligible is an indicator for whether the student was ever eligible for free or reduced-price lunch during the experiment. See notes to Tables V and VIII for regression specifications and other variable definitions. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE XII
Impacts of Teacher Experience: Small vs. Large Classes

Dependent Variable:	Test Score	Wage Earnings	Test Score	Wage Earnings
	(%)	($)	(%)	($)
A. Small Classes	(1)	(2)	(3)	(4)
Teacher with >10 years of experience	2.91	405.4	0.187	-2244
	(3.15)	(1326)	(3.54)	(1750)
Observations	1,690	1,817	854	1,047
B. Large Classes				
Teacher with >10 years of experience	2.95	1471	0.695	-540.2
	(1.68)	(714.2)	(1.47)	(702.3)
Observations	3,911	4,188	3,416	3,862
Entry grade	KG	KG	Grade ≥1	Grade ≥1

Notes: This table replicates the specifications of Columns 1-4 of Table VI. Panel A includes students assigned to small classes upon entry; Panel B includes those assigned to large classes. See notes to Table VI for regression specifications and variable definitions. Standard errors, reported in parentheses, are clustered by school.

APPENDIX TABLE XIII

Effects of Class Quality on Earnings: Instrumental Variable Estimates

Dependent Variable:	Wage Earnings ($)				
A. All Entrants	(1)	(2)	(3)	(4)	(5)
Entry-grade test percentile	82.21	96.39			90.04
	(23.63)	(31.16)			(8.65)
B. KG Entrants					
KG test percentile	78.71	89.28	74.85	80.33	93.79
	(35.09)	(39.70)	(26.50)	(26.40)	(9.56)
Estimation method	Leave-Out Mean	Split-Sample	LIML	2SLS	OLS

Notes: The effects of class quality on test scores and earnings reported in Columns 1 and 2 of Table VIII can be combined to produce a reduced-form IV estimate of the earnings effect associated with an increase in test scores: $50.61/0.662=$76.48. Including only those observations with both test score and wages in the data changes the coefficient slightly to $82.21 in Column 1. To test the robustness of our leave-out mean estimator, we report three alternative IV estimates of the impact of test scores on earnings, controlling for school-by-entry-grade fixed effects and the demographic controls used in Column 2 of Table VIII. Column 2 instead uses a split-sample definition of peer scores, in which we randomly split classes into two groups and proxy for class quality using peer scores in the other group. Column 3 estimates the model using limited information maximum likelihood, using kindergarten class dummies as instruments. Column 4 instruments for test score with classroom dummies using two-stage least squares. Column 5 reports an OLS estimate of the correlation between test scores and earnings as a reference. Panel A pools all entry grades and reports estimates for three of the specifications. Instrumenting with entry class dummies is ill defined in later grades as it would require defining class quality based purely on the new entrants, who constitute a small fraction of each class. Panel B replicates the specifications in A using the subsample of kindergarten entrants. The leave-out mean IV estimator in Panel B Column 1 coincides with the jackknife IV of Angrist, Imbens, and Krueger (1995), when we use only KG entrants. When we pool entry grades, our leave-out mean estimator differs from jackknife IV because we form the leave-out mean measure of class quality using all peers (including previous entrants), not just those who entered in the current entry grade. Standard errors, reported in parentheses, are clustered by school in all columns except 3.

APPENDIX TABLE XIV
Effects of Class Quality on Components of Non-Cognitive Measures

A. Cross-Sectional Correlations

Dependent Variable:	Wage Earnings							
	Grade 4 Non-Cognitive Measure				Grade 8 Non-Cognitive Measure			
	($)	($)	($)	($)	($)	($)	($)	($)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Effort	79.91				127.8			
	(16.52)				(15.43)			
Initiative		57.06				92.29		
		(18.08)				(16.86)		
Value			61.47				115.5	
			(16.97)				(19.68)	
Participation				37.26				66.34
				(18.68)				(16.38)

B. Class Quality Impacts

	Grade 4 Non-Cognitive Measure				Grade 8 Non-Cognitive Measure			
Dependent Variable:	Effort	Initiative	Value	Particip	Effort	Initiative	Value	Particip
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Class quality	0.151	0.165	0.095	0.120	0.141	0.070	0.124	0.178
(peer scores)	(0.066)	(0.070)	(0.071)	(0.082)	(0.071)	(0.056)	(0.069)	(0.066)

Notes: This table decomposes relationships described in Table IX into the four constituent components of non-cognitive skill. These four non-cognitive measures are constructed from a series of questions asked of the student’s teacher(s) and are intended to measure, respectively: student effort in class, initiative, whether a student perceives school/class as "valuable", and participatory behavior. The measures were reported twice, once by the student’s regular 4th grade teacher (Columns 1-4) and the second time the scores are the average of the reports by the 8th grade math and English teachers (Columns 5-8). Each of the four variables is scaled as a within-sample percentile rank. Panel A replicates Column 1 of Table IX, using only one of the four non-cognitive measures as a covariate in each regression. Panel B replicates Column 5 of Table IX, using one of the four non-cognitive measures as the dependent variable in each regression. See notes to Table IX for regression specifications and other variable definitions. Standard errors, reported in parentheses, are clustered by school.