GERALD R. FORD SCHOOL OF PUBLIC POLICY
EDUCATION POLICY INITIATIVE
UNIVERSITY OF MICHIGAN

Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion

Susan M. Dynarski, *University of Michigan*
Joshua M. Hyman, *University of Connecticut*
Diane Whitmore Schanzenbach, *Northwestern University*

Education Policy Initiative
Gerald R. Ford School of Public Policy
735 S. State Street
Ann Arbor, Michigan 48109

EPI Working Paper 04-2013 | July, 2013

EXPERIMENTAL EVIDENCE ON THE EFFECT OF CHILDHOOD INVESTMENTS ON POSTSECONDARY ATTAINMENT AND DEGREE COMPLETION

Susan Dynarski
Joshua M. Hyman
Diane Whitmore Schanzenbach

October 2011, Revised July 2013

Originally posted as NBER Working Paper #17533

We thank Jayne Zaharias-Boyd of HEROS and the Tennessee Department of Education for allowing the match between the STAR and National Student Clearinghouse data. The Education Research Section at Princeton University generously covered the cost of this match. Monica Bhatt, David Deming and Nathaniel Schwartz provided excellent research assistance. We benefitted from discussions at Cornell, the Federal Reserve Bank of Atlanta, the Swedish Institute for Labour Market Evaluation, University of California at Davis, University of Michigan, Vanderbilt, Yale and the 2012 Rome conference on Improving Education Accountability and Evaluation. The views expressed herein are those of the authors and do not necessarily reflect the views of the Education Policy Initiative.

Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion
Susan Dynarski, Joshua M. Hyman, and Diane Whitmore Schanzenbach
October 2011, Revised July 2013

ABSTRACT

This paper examines the effect of early childhood investments on college enrollment and degree completion. We use the random assignment in the Project STAR experiment to estimate the effect of smaller classes in primary school on college entry, college choice, and degree completion. We improve on existing work in this area with unusually detailed data on college enrollment spells and the previously unexplored outcome of college degree completion. We find that assignment to a small class increases the probability of attending college by 2.7 percentage points, with effects more than twice as large among blacks. Among students enrolled in the poorest third of schools, the effect is 7.3 percentage points. Smaller classes increase the likelihood of earning a college degree by 1.6 percentage points and shift students towards high-earning fields such as STEM (science, technology, engineering and mathematics), business and economics. We find that test score effects at the time of the experiment are an excellent predictor of long-term improvements in postsecondary outcomes.

Susan Dynarski
University of Michigan
Weill Hall
735 South State Street
Ann Arbor, MI 48109-3091
and NBER
dynarski@umich.edu

Joshua M. Hyman
University of Michigan
Weill Hall
735 South State Street
Ann Arbor, MI 48109-3091
http://www-personal.umich.edu/~jmhyman/
jmhyman@umich.edu

Diane Whitmore Schanzenbach
School of Education and Social Policy
Northwestern University
Annenberg Hall, Room 205
2120 Campus Drive
Evanston, IL 60208
and NBER
dws@northwestern.edu

Education is intended to pay off over a lifetime. Economists conceive of education as a form of "human capital," requiring costly investments in the present but promising a stream of returns in the future. Looking backward at a number of education interventions (e.g., Head Start, compulsory schooling), researchers have identified causal links between these policies and long-term outcomes such as adult educational attainment, employment, earnings, health and civic engagement (Ludwig & Miller, 2007; Deming, 2009; Angrist & Krueger, 1991; Dee, 2004; Lleras-Muney, 2005). But decision-makers attempting to gauge the effectiveness of current education inputs, policies and practices in the present cannot wait decades for these long-term effects to emerge. They therefore rely upon short-term outcomes – primarily standardized test scores – as their yardstick of success.

A critical question is the extent to which short-term improvements in test scores translate into long-term improvements in well-being. Puzzling results from several evaluations make this a salient question. Three small-scale, intensive preschool experiments produced large effects on contemporaneous test scores that quickly faded (Schweinhart, et al., 2005; Anderson, 2008). Quasi-experimental evaluations of Head Start, a preschool program for poor children, reveal a similar pattern, with test score effects gone by middle school. In each of these studies, treatment effects have re-emerged in adulthood as increased educational attainment, enhanced labor market attachment, and reduced crime (Deming, 2009; Garces, Thomas, & Currie, 2002; Ludwig & Miller, 2007). Further, several recent papers have shown large impacts of charter schools on test scores of disadvantaged children (Abdulkadiroglu, et al., 2011; Angrist, et al., 2012; Dobbie & Fryer, 2011). A critical question is whether these effects on test scores will persist in the form of long-term enhancements to human capital and well-being.

We examine the effect of smaller classes on educational attainment in adulthood, including college attendance, degree completion and field of study. We exploit random variation in class size in the early grades of elementary school created by the Tennessee Student/Teacher Achievement Ratio (STAR) Experiment. Participants in the STAR experiment are now in their thirties, an age at which it is plausible to measure completed education. Our postsecondary outcome data is obtained from the National Student Clearinghouse (NSC), a national database that covers approximately 90 percent of students enrolled in colleges in the U.S.

We find that being assigned to a small class increases the rate of postsecondary attendance by 2.7 percentage points. The effects are considerably higher among populations with traditionally low rates of postsecondary attainment. For Black students and students eligible for free lunch the effects are 5.8 and 4.4 percentage points, respectively. At elementary schools with the greatest concentration of poverty, measured using the fraction of students receiving a subsidized lunch, smaller classes increase the rate of postsecondary attendance by 7.3 percentage points. We further find that being assigned to a small class increases the probability of earning a college degree by 1.6 percentage points. Smaller classes shift students toward earning degrees in high-earning fields such as science, technology, engineering and mathematics (STEM), business and economics.

Our results shed light on the relationship between the short-and long-term effects of educational interventions. The short-term effect of small classes on test scores, it turns out, is an excellent predictor of its long-term effect on adult outcomes. We show this by adding K-3 test scores to our identifying equation; the coefficient on the class size dummy drops to zero. The coefficient on the interaction of class size and test scores is also zero, indicating that the scores of children in small classes are no less (or more) predictive of adult educational attainment than those of children in the regular classes.

Our analysis identifies the effect of manipulating a single policy-relevant educational input on adult educational attainment. By contrast, the early-childhood interventions for which researchers have identified lifetime effects (e.g., Head Start, Abecederian) are multi-pronged, including home visits, parental coaching and vaccinations in addition to time in a preschool classroom. We cannot distinguish which dimensions of these treatments generate short-term effects on test scores, and whether they differ from the dimensions that generate long-term effects on adult well-being. The effective dimensions of the treatment are also ambiguous in the recent literature on classroom and teacher effects. For example, Chetty et al. (2011) show very large effects of kindergarten classroom assignment on adult well-being. In those estimates, the variation in classroom quality that produces significant variation in adult outcomes excludes class size but includes anything else that varies at the classroom level, including teacher quality and peer quality, both of which are extremely difficult to manipulate with policy. By contrast, the effects we measure in this paper, both short-term and long-term, can be attributed to a well-defined and replicable intervention: reduced class size.

THE TENNESSEE STAR EXPERIMENT

The Tennessee Student/Teacher Achievement Ratio (STAR) Experiment randomly assigned class sizes to children in kindergarten through third grade. The experiment was initiated in the 1985-86 school year, when participants were in kindergarten. A total of 79 schools in 42 school districts participated, with over-sampling of urban schools. An eventual 11,571 students were involved in the experiment. The sample is 60 percent white and the balance African American. About 60 percent of students were eligible for subsidized lunch during the experiment. The experiment is described in greater detail elsewhere (Word, et al., 1990; Folger & Breda, 1989; Finn & Achilles, 1990; Krueger, 1999; and Achilles, 1999.)

Children in the STAR experiment were assigned to either a small class (target size of 13 to 17 students) or regular class (22 to 25 students).[1] Students who entered a participating school after kindergarten were randomly assigned during those entry waves to a regular or small class. Teachers were also randomly assigned to small or regular classes. All randomization occurred within schools.

Documentation of initial random assignment in STAR is incomplete (Krueger, 1999). Krueger (1999) examines records from 18 STAR schools for which assignment records are available. He finds that, as of entry into STAR, 99.7 percent of students were

[1] A third arm of the experiment assigned a full-time teacher's aide to regular classes. Previous research has shown no difference in outcomes between the regular-sized classes with and without an aide. We follow the previous literature in pooling students from both types of regular classes into a single control group. The results are substantively unchanged if we include an indicator variable for the presence of a full-time teacher's aide.

enrolled in the experimental arm to which they were initially assigned. Krueger's approach, and that of the subsequent literature, is to assume that the class type in which a student is first enrolled is the class type to which she was assigned. We follow that convention in our analysis.

Numerous papers have tested, and generally validated, the randomization in STAR (Krueger, 1999). There are no baseline outcome data (e.g., a pre-test) available for the STAR sample. On the handful of covariates available in the STAR data (free lunch eligibility, race, sex), the arms of the experiment appear balanced at baseline (see Table 1 for a replication of these results). Recent work by Chetty et al. (2011) shows that the STAR entry waves were balanced at baseline on a detailed set of characteristics (e.g., family income, home ownership) contained in the income tax returns of the STAR subjects' parents.

PREVIOUS RESEARCH ON THE LONG-TERM EFFECTS OF SMALL CLASSES

A substantial body of research has examined the effect of Project STAR on short-and medium-run outcomes. We do not comprehensively discuss this literature but instead summarize the pattern of findings. These papers show that students assigned to a small class experience contemporaneous test score gains of about a fifth of a standard deviation. These test score results diminish after the experiment ends in third grade.[2] There is evidence of lasting effects on other dimensions. Krueger and Whitmore (2001) show that students assigned to small classes are more likely to take the ACT and SAT, required for admission to most four-year colleges. Schanzenbach (2006) reports that smaller classes reduce the rate of teen pregnancy among female participants by about a third. In addition, Fredriksson, Ockert, and Oosterbeek (2013) find positive long-term impacts of reduced class size in grades 4-6 in Sweden on educational attainment and wages.

The paper most closely related to our own examined the impact of Project

[2] Cascio and Staiger (2012) show that fade-out of test-score effects is, at least in some settings, a statistical artifact of methods used by analysts to normalize scores within and across grades. However, they specifically note that the sharp drop in estimated effects that occurs after the end of the STAR experiment cannot be explained in this way.

STAR on adult outcomes using the income tax records of STAR participants and their parents (Chetty et al., 2011). That paper emphasizes the differential long-term impacts of being randomly assigned to classrooms of different “quality” levels stemming from higher-quality teachers and/or classmates, after accounting for class size. Chetty et al. (2011) document the sizeable long-term payoff to having a high quality classroom, though recognize that this cannot be directly manipulated by public policy. By contrast, we focus on the long-term impacts of randomly assigned class size, which is an easily measured input that can be manipulated by policy.

EMPIRICAL STRATEGY

The experimental nature of Project STAR motivates the use of a straightforward empirical specification. We compare outcomes of students randomly assigned to small and regular classes by estimating the following equation using Ordinary Least Squares:

$$y_{isg} = \beta_0 + \beta_1 SMALL_{is} + \beta_2 X_{is} + \beta_{sg} + \varepsilon_{isg} \,, \qquad (1)$$

where y_{isg} represents a postsecondary schooling outcome of student i, who entered the STAR experiment in school s and in grade g. X is a vector of covariates including sex, race and free lunch status (an indicator for whether the student ever received free or reduced price lunch during the experiment), included to increase precision. β_{sg} is a set of school-by-entry-grade fixed effects. We include these because students who entered STAR schools after kindergarten were randomly assigned at that time to small or regular classes. The variable of interest is $SMALL_{is}$, an indicator set to one if student i was assigned to a small class upon entering the experiment. The omitted group to which small classes are compared is regular classes (with or without a teacher’s aide). We cluster standard errors by school, the most conservative approach. Standard errors are about ten percent smaller if we cluster at the level of school-by-wave.

DATA

We use the original data from the STAR experiment, which includes information on the type of class in which a student is enrolled, basic demographics (race, poverty status, sex), school identifiers, and standardized test scores. These data also include the name and date of birth of the student, which we use to match to data on postsecondary attainment and completion.

Data on postsecondary outcomes for the STAR sample come from the National Student Clearinghouse (NSC). NSC is a non-profit organization that was founded to assist student loan companies in validating students' college enrollment. Borrowers can defer payments on most student loans while in college, which makes lenders quite interested in tracking enrollment. Colleges submit enrollment data to NSC several times each academic year, reporting whether a student is enrolled, at what school, and at what intensity (e.g., part-time or full-time). NSC also records degree completion and the field in which the degree is earned. States and school districts use NSC data to track the educational attainment of their high school graduates (Roderick, Nagaoka, & Allensworth, 2006). Recent academic papers making use of NSC data include Deming et al. (2011) and Bettinger et al. (2012).

With the permission of the Project STAR researchers and the state of Tennessee, we submitted the STAR sample to the NSC in 2006 and again in 2010.[3] The STAR sample was scheduled to graduate high school in 1998. We therefore capture college enrollment and degree completion for twelve years after on-time high-school graduation, when the STAR sample is about 30 years old.

The NSC matches individuals to its data using name and date of birth. If birth date is missing, the NSC attempts to match on name alone. Some students in the STAR sample are missing identifying information used in the NSC match: 12 percent have incomplete name or birthdate. In our data, a student that attends college but fails to produce a match in the NSC database is indistinguishable from a student who did not attend college. If the absence of these identifiers is correlated with the treatment, then our estimates may be biased. To determine whether identifiers are missing at a differential rate across treatment groups, we estimate equation (1) replacing y_{isg} with an indicator variable equaling one if a student has a missing name or date of birth. We find a precisely estimated zero for β_1 (=-0.008, SE=0.008) indicating that the probability of missing identifying information is uncorrelated with initial assignment. In the concluding

[3] In 2006, the NSC used social security number as well as name and date of birth in its matches. As of 2010, NSC had ceased to use social security number for its matches.

section of the paper, we present the results of a second test exploring the possible bias in our main result associated with missing identifiers.

Not all schools participate in NSC; the company estimates they currently capture about 93 percent of undergraduate enrollment nationwide. During the late 1990s, when the STAR subjects would have been graduating from high school, the NSC included colleges enrolling about 80% of undergraduates in Tennessee (Dynarski, Hemelt, & Hyman, 2012).[4] Since we miss about 20% of undergraduate enrollment using the NSC data, we expect that we will underestimate the college attendance rate of the STAR sample by about a fifth. The NSC data indicate that 39.4 percent of the STAR sample had attended college by age 30. Among those born in Tennessee in the same years as the STAR sample, the attendance rate is 52.8 percent in the 2005 American Community Survey (Ruggles, et al., 2010).[5] Our NSC estimate of college attendance is therefore, as expected, about four-fifths of the magnitude of the ACS estimate.

In the NSC, we find that 15.1 percent of the STAR sample has earned a college degree. This is substantially lower than the corresponding rate we calculate from the 2005 American Community Survey (29.3 percent). Not all of the colleges that report enrollment to the NSC report degree receipt, and this explains at least part of the discrepancy.[6]

The exclusion of some colleges from NSC will induce measurement error in the dependent variable. If this error is not correlated with treatment (i.e., classical measurement error) then the true effect of class size on college enrollment will be

[4] Dynarski et al. (2012) calculate this rate by dividing undergraduate enrollment at Tennessee colleges included in NSC as of 1998 by enrollment at all Tennessee colleges in 1998. The list of colleges participating in the NSC and the year that they joined is accessible on the NSC website. Enrollment data are from the Integrated Postsecondary Education Data System (IPEDS), a federally-generated database that lists every college, university and technical or vocational school that participates in the federal financial aid programs (about 6,700 institutions nationwide) (National Center For Education Statistics, 2010).

[5] We re-weight the Tennessee-born in the ACS data to match the racial composition of the STAR sample, which was disproportionately black.

[6] Using IPEDS, we calculate that 70% of undergraduate degrees are conferred by institutions that, according to the NSC website, report degrees to NSC. Dynarski et al. (2012) also find lower degree coverage in the NSC relative to enrollment coverage.

larger than our observed effect by the proportion of enrollment that is missed (approximately 20 percent).[7] This is because the true treatment effect is the sum of the observed treatment effect and the treatment effect of the unobserved college attenders (Bound, Brown, & Mathiowetz, 2001). However, if the measurement error in college attendance is correlated with assignment to treatment then our effect could be either downward or upward biased. This would be the case, for example, if colleges attended by marginal students are disproportionately undercounted by NSC.

To determine whether the NSC systematically misses certain types of schools, we compare the schools that participate in NSC with those in IPEDS. Along all measures we examined (i.e., sector, racial composition, selectivity), the NSC colleges are similar to the universe of IPEDS colleges, with a single exception: NSC tends to exclude for-profit institutions.[8] These are primarily trade schools such as automotive, technology, business, nursing, culinary arts and beauty schools. If small classes tend to induce into such schools those students who would not otherwise attend college, we will underestimate the effect of small classes on college attendance. If on the other hand small classes induce students out of such schools into colleges that we tend to observe, such as community colleges, then our estimates will be upward biased. In the concluding section of our paper, we conduct a back-of-the-envelope exercise to bound the possible upward bias that could be due to this phenomenon.

RESULTS

In this section, we examine the effect of assignment to a small class on a set of postsecondary outcomes: college entry, the timing of college entry, college choice, degree receipt and field of degree.

College Entry

In Table 2, we estimate the effect of assignment to a small class on the probability of

---

[7] This is true in terms of percentage points. The percent increase in college attendance would remain unchanged.

[8] The conclusion is the same when we weight coverage by the number of degrees conferred rather than by undergraduate enrollment.

college entry by age 30. The effect is close to three percentage points (Column 1, 2.8 percentage points), which is an impact of approximately 7 percent relative to the control mean of 38.5 percent (control means are italicized in the tables). This estimate is statistically significant, with a standard error of about one percentage point. Including covariates does not alter the estimate, as is expected with random assignment. For the balance of the paper we report results that include covariates, since they are slightly more precise.

Splitting the sample by race reveals that the effects are concentrated among Blacks (5.8 points relative to a mean of 30.8 percent) and those eligible for free or reduced-price lunch (4.4 points relative to a mean of 27.2 percent). The effects are twice as large for boys (3.2 points relative to a mean of 32.4 percent) than girls (1.6 points relative to a mean of 45.5 percent). Breaking the effects down yet more finely shows that the effects are largest for Black females (7.2 points, standard error of 3.5), with no effect on white females (1.3 points, standard error of 2.3). The effects for Black and white males are indistinguishable (3.1 and 4.4 points, respectively; standard error of 1.8 and 2.4 points).

One caveat to consider when examining results by race and gender is that the probability of enrolling in a college not in the NSC could be correlated with race-gender, which could cause bias in the estimates. Dynarski et al. (2012) show that NSC coverage is similar by sex, but is lower for Black students than white students. To examine this issue for a population similar to the STAR sample of students, we examine the share of first-time college students in Tennessee in 1998 in IPEDS by race and sex attending for-profits (which tend not to appear in NSC) and attending any type of college. We find that black and female students tend to enroll in higher proportions in for-profit colleges. This suggests that part of the large treatment effect for black females could be due to these students being induced from non-NSC colleges to those that participate in NSC.

Our results by student demographics indicate that there is substantial heterogeneity by race and income in the effect of class size. However, policy decisions regarding staffing levels and class size tend to be set at the school level rather than the student level. School-level characteristics, rather than student-level characteristics, may therefore be the more policy-relevant dimension along which to measure heterogeneity in effects. In order to capture this policy-relevant variation in effects, we divide the STAR schools into three groups: those with low, medium and high levels of poverty, which we proxy with the share of children eligible for a subsidized lunch. We sort students by this share, and construct the groups such that the number of students in each group is nearly identical (see Appendix Table 1). Note that the STAR sample was disproportionately poor and urban, so even the schools with the lowest levels of poverty are relatively disadvantaged.

When we estimate Equation (1) separately for these three groups of schools, we find that the treatment effect is concentrated in the poorest schools. At schools with low to medium concentrations of poverty, the estimated effect of class size on postsecondary attainment is indistinguishable from zero (Table 2, Columns 7 and 8). But the estimated effect is 7.3 percentage points in the poorest schools. This is a 28 percent increase relative to the control mean in these schools. A test of the equality of the coefficients for the poorest schools versus the combined bottom two terciles is strongly rejected (p-value of 0.008, Column 11).

Inequality in postsecondary education has increased in recent decades, with the gap in attendance between those born into lower-income and higher-income families expanding (Belley & Lochner, 2007; Bailey & Dynarski, 2011). The pattern of effects described above will tend to decrease gaps in postsecondary attainment. Figure I shows this graphically. On the top is depicted the gap in college attendance between blacks and whites in regular classes (left) and in small classes (right). The black-white gap is about half as large in small classes (7.7 percentage points) as it is in regular classes (12.4 percentage points). The drastic reduction in the race gap in college attendance is driven by females, for whom the race gap virtually disappears in small classes (results not shown).

In the control group, students who were eligible for free or reduced-price lunch are 29.1 percentage points less likely to attend college than their higher-income classmates. The gap is slightly smaller in the treatment group (25.7 percentage points). Finally, we compare the effect of small classes on the gap in postsecondary outcomes between schools with high and moderate levels of poverty. Among students in regular-sized classes, the gap in postsecondary attendance is 18.1 percentage points. Among students in small classes, the gap is nearly halved, to 9.8 percentage points.

Class size could plausibly affect the intensity with which a student enrolls in college, in addition to the decision to enroll at all. The overall impact on the intensity of enrollment is theoretically ambiguous: students induced into college by smaller classes may be more likely to enroll part-time than other students, while treatment could induce those who would have otherwise enrolled part-time to instead enroll full-time. In the control group, about three-quarters of college entrants (ever) attend college full-time, while a quarter never do (Table 2, second row). When we re-estimate Equation (1) with these two variables as dependent variables, we find that the effect on entry is evenly divided between part-time and full-time enrollment. While the standard errors preclude any firm conclusions, these results suggest that the marginal college student is more likely than the inframarginal student to attend college exclusively on a part-time basis.

Timing of College Attendance

Class size could plausibly affect the timing of postsecondary attendance. The net effect is theoretically ambiguous. Smaller classes may lead students who would otherwise have attended college to advance through high school more rapidly, enter college sooner after graduation, and move through college more quickly. On the other hand, students induced into college by smaller classes may enter and move through college at a slower pace than their inframarginal peers.

We first estimate the effect of class size upon "on-time enrollment," which we define as entering college by fall of 1999, or about 18 months after the STAR cohort is scheduled to have graduated high school. This variable captures the pace at which students complete high school, how quickly they enter college, and whether they attend college at all. By this measure, 27.4 percent of the control group has enrolled on-time, or about three-quarters of the 38.5 percent who ever attend college (Table 2). Assignment to a small class increases the likelihood of entering college on time by 2.4

percentage points. Among those students enrolled in the poorest third of schools, the effect is 4.7 points, a 29 percent increase relative to this group's control mean of 16 percent. These results suggest that students in smaller classes are no less likely to start college on time than control students: 72 percent of the treatment-group students who attend college do so on time, while among the control group the share of attendance that is on-time is 71 percent.

We next look at the year-by-year evolution of the effect of class size on postsecondary attainment. For each year, we plot the share of students who have ever attended college, separately for the treatment and control group (Figure II, top panel). We also plot the treatment-control difference, along with its 95% confidence interval (Figure II, bottom panel). The fraction of the sample that has ever attended college rises from under 5 percent in 1997 to over 20 percent in 1998 (when students are 18). The rate rises slowly through age 30, when the share of the sample with any college experience reaches nearly 40 percent. The difference between the two groups reaches about three points by age 19 and remains at that level through age 30.[9] When we examine the shares of students who are currently enrolled in college (Figure III) we see that the treatment group is more likely to be enrolled in college at every point in time, peaking at around 25 percent in 1999. Plausibly, smaller classes could have sped up college enrollment and completion, and the control group could eventually have caught up with the treatment group in its rate of college attendance. This is not what we see, however. The effect is always positive, and is largest right after high school, when the sample is 18 to 19 years old.[10]

College Choice

By boosting academic preparation, smaller classes in primary school may induce

[9] To obtain the figures, we replace the small-class indicator variable in our identifying equation with a full set of its interactions with year fixed effects. The coefficients on these interactions and their confidence intervals are plotted in the bottom panel. In the top panel, we add these interactions to the year-specific control means.

[10] This pattern of findings sheds light on the difference between our findings and those of Chetty et al. (2011). We can reconcile our findings with Chetty et al. (2011) if we censor the NSC data so that they exclude the same enrollment spells that are unobserved in their data, see Appendix Table 2.

students to alter their college choices. For example, those who would have otherwise attended a two-year community college may instead choose to attend a four-year institution. Bowen, Chingos, and McPherson (2009) suggest that attending higher quality colleges (which provide more inputs, including better peers) is a mechanism through which students could increase their rate of degree completion.

In Table 3, we examine the effect of class size on college choice. Across the entire sample, we find little evidence that exposure to smaller classes shifts students toward higher-quality schools. The treatment effect is concentrated on attendance at two-year institutions. While 22 percent of the control group starts college at a two-year school, the rate is 2.5 percentage points higher in the treatment group (with a standard error of 0.9 percentage points). The effect is 6.3 percentage points among students in the poorest third of schools. We find positive but imprecise effects on the probability of ever attending a four-year college, attending college outside Tennessee, or attending a selective college.[11]

Persistence and Degree Completion

While college entry has been on the rise in recent decades, the share of college entrants completing a degree is flat or declining (Bound, Lovenheim, & Turner, 2010). About half of college entrants never earn a degree. A key concern is that marginal students attending college may drop out quickly, in which case the attendance effects discussed above would overestimate the effect of class size on social welfare.

We explore this issue by examining the effect of small classes on the number of semesters that students attend college, as well as on the probability that they complete a college degree. Overall, the number of semesters attempted (including zeroes) is quite low: the control group attempts an average of three semesters by age 30. Among those in the control group with any college experience, the average number of semesters attempted is eight.

The treatment group spends 0.22 more semesters in college than the control

[11] We measure selectivity using Barron's quality categories. Using an index that includes multiple proxies for quality such as the acceptance rate, tuition, and the average ACT/SAT score of entering students provides similar results.

group (Figure IV, top; Table 4). The effects are somewhat larger among students in the poorest schools (coefficient of 0.32), though the effect is imprecisely estimated and the difference across terciles is less stark than with the college entry effects. The size of these effects is comparable to treatment effects found in the Opening Doors demonstration, which gave short-term rewards to community college students for achieving certain enrollment and grade thresholds (Barrow, et al., 2009).

Assignment to a small class increases the likelihood of completing a college degree by 1.6 percentage points (Table 4); the result is statistically significant at the 10 percent level. When we examine effects separately by highest degree earned, we find that the 1.6 percentage point effect is driven evenly by increases in 2-year (associates) and 4-year (bachelors) degree receipt (0.7 and 0.9 percentage points, respectively). When we turn to the timing of degree completion, we see that there is a positive treatment effect at every age. The difference is largest between age 22 and 23 (Figure IV, Panel C). Students assigned to small classes during childhood continue to outpace their peers in their rate of degree completion well into their late twenties. This may explain why Chetty et al. (2011) do not find an effect of small classes on earnings, which they observe at age 27. Members of the treatment group are still attending and completing college at this age, and so have likely not yet spent enough time in the labor market for their increased education to offset experience forgone while in college.

Field of Degree

The earnings of college graduates vary considerably by field. In particular, those who study science, technology, engineering and mathematics (STEM), as well as business and economics, enjoy higher returns than other college graduates (Arcidiacono, 2004; Hamermesh & Donald, 2008). In this section we examine whether class size affects the field in which a student completes a degree.[12]

We divide degrees into three categories: 1) STEM fields; business and economics concentrations; and all others.[13] Students can earn more than one degree (e.g., an AA

[12] Field of study is available only for students who complete a degree; we are therefore unable to examine the field of study for non-completers.

[13] We follow a degree-coding scheme defined by the National Science Foundation

and a BA); we code them as having a STEM degree if any degree falls in this category, and as having a business or economics degree if any degree falls in this category and they have not earned a STEM degree. In practice, very few students earn both a STEM and business or economics degree.

Assignment to a small class shifts the composition of degrees toward STEM, business and economics. While 1.9 (2.6) percent of the control group earns a degree in a STEM (business or economics) field, the rate is 2.4 (3.3) in the treatment group (Table 4). However, these estimates are imprecisely estimated. In order to increase precision and to group fields by whether or not they are high-paying, we combine the STEM, business and economics fields into one category. Assignment to a small class increases degree receipt in these high-paying fields by 1.3 percentage points. This difference is statistically significant at the 5 percent level, with a standard error of 0.6 percentage points. There is no difference in the rate at which students receive degrees in other fields.

These results are consistent with two scenarios: (1) those induced into completing a degree tend to concentrate in STEM, business and economics or (2) inframarginal degree completers are shifted toward STEM, business and economics. While we cannot conclusively identify those who are and are not on the margin of completing a degree, our analysis by school-level poverty tercile (Table 4, Columns 2 and 3) suggests that the second scenario is at work. The effect of small classes on graduating in a STEM, business or economics degree is 1.9 percentage points (standard error of 0.8 points) among the less poor schools where students are more likely to be inframarginal degree completers. The effect is zero among the poorest third of schools, where students are more likely to be induced into completing a degree. These effects are statistically different from one another at the 10 percent level.

Testing for Sources of Heterogeneity in Effects

(National Science Foundation, 2011). We apply this scheme to two text fields included in NSC: degree title (e.g., “associates” or “bachelor of science”) and college major (e.g., “biology”). A small number of students who receive a degree are missing both degree title and college major, and are excluded from this analysis.

One interpretation of these results is that the groups with the lowest control means are most sensitive to class size. An alternative interpretation, however, is that the groups that display the largest response are actually exposed to a more intense dosage of the treatment. All of our estimates so far have been of the effect of the intention to treat (ITT), which is attenuated toward zero when there is crossover and noncompliance. The groups that show the largest ITT effects may have received larger dosages of the treatment, in the form of particularly small classes or more years spent in a small class. Krueger and Whitmore (2002) show that disadvantaged students in the treatment group are not systematically assigned to the smallest of the small classes. Here we examine whether they are exposed to more years in a small class.

We generate subgroup estimates of the effect of assignment to a small class on years spent in a small class. To do so, we instrument for years actually spent in a small class with years potentially spent in a small class. Potential years in a small class is the product of assignment to a small class and the number of years the student could be enrolled in a small class, based on year of entry into STAR. For example, a student who entered STAR in kindergarten could spend as many as four years in a small class, while a child who entered in third grade could spend only one.[14]

We estimate the following equations:

$$YEARS_{is} = \delta_0 + \delta_1 Z_{is} + \delta_{sg} + \psi_{isg} \qquad (2)$$

$$COLL_{isg} = \alpha_0 + \alpha_1 YEARS_{is} + \alpha_{sg} + \varepsilon_{isg}\ , \qquad (3)$$

where $COLL_{isg}$ is an indicator variable for whether student i, who entered the STAR experiment in school s and in grade g, ever enrolls in college. YEARS is the number of years the student spends in a small class. Z is the potential number of years a student could attend a small class multiplied by an indicator for whether the student was assigned to a small class. School-by-entry-grade fixed effects are included in each equation. We estimate these equations separately by subgroup.

Table 5 reports the estimates of the first stage equation, the reduced-form

[14] Abdulkadiroglu et al. (2011) and Hoxby and Murarka (2009) use a similar approach when they instrument for years spent in a charter school with potential years spent in a charter school, where potential years is a function of winning a charter lottery and the grade of application.

intention-to-treat model (ITT) and the two-stage least squares model (2SLS). The first-stage results in column (1) measures compliance, reporting the number of years actually spent in a small class for each year assigned to a small class. Overall, for each year of potential small-class attendance, students on average attend 0.64 years in a small class. The compliance rate is consistently smaller for the groups for whom we have estimated the largest effects of ITT. This is likely driven by higher mobility among black and poor students. The 2SLS estimates (Column 3) indicate that each year spent in a small class increases college attendance rates by one percentage point for the entire sample, but by 2.8 points for students attending the poorest schools, 2.4 points for black students, and 1.6 points for poor students. These results indicate that students who are black, poor, or attend high-poverty schools benefit more from a year spent in a small class than do their peers.

Do Short-Term Effects Predict Long-Term Effects?

We have shown that random assignment to small classes increases college entry and degree completion and shifts students toward high-paying majors. Could these effects have been predicted by the short-term effects of STAR on test scores? That is, are the effects measured at the time of the experiment predictive of the program's long-term effects?

A back-of-the-envelope prediction would combine the experiment's effect on scores with information from some other data source on the relationship between scores and postsecondary attainment. We now make such an informed guess about the long-term effects of STAR, then compare our guess with the paper's findings.

The guess requires information about the relationship between standardized scores in childhood and adult educational attainment, ideally for a cohort born around the same time as the STAR subjects. The NLSY79 Mother-Child Supplement contains longitudinal data on the children of the women of the National Longitudinal Survey of Youth. These children were born at roughly the same time as the STAR cohort. The children of the NLSY (CNLSY) were tested every other year, including between the ages of six and nine (the ages of the STAR subjects while the experiment was underway).

Postsecondary attainment is also recorded in CNLSY.

In CNLSY a standard deviation increase in childhood test scores is associated with a 16 percentage-point increase in the probability of attending college.[15] Assignment to a small class in STAR increases the average of K-3 scores by 0.17 standard deviations. Under the assumption that the relationship between scores and attainment is the same for the STAR and NLSY79 children, a reasonable prediction of the effect of STAR on the probability of college attendance is 2.72 percentage points (=0.17*16). This back-of-the-envelope calculation is nearly identical to the 2.7 point estimate we obtained in our regression analysis, indicating that the contemporaneous effect of STAR on scores is an excellent predictor of its effect on adult educational attainment.

Another way to approach this question is to examine whether the estimated effect of small classes on postsecondary attainment disappears when we control for K-3 test scores. This is an informal test of whether class size affects postsecondary attainment through any channel other than test scores. This sort of informal test is often used when checking whether an instrument (e.g., assigned class size) affects the outcome of interest (e.g., postsecondary attainment) through any channel other than the endogenous regressor (e.g., test scores). We first re-estimate Equation (1) and report the main result in column 1 of Table 6. We then add to this regression a student's test scores and the interaction of the test scores and assignment to a small class. The interaction allows the relationship between test scores and postsecondary attainment to differ between small and regular classes:

$$Coll_{isg} = \beta_0 + \beta_1 SMALL_{is} + \beta_2 TEST_{is} + \beta_3 SMALL_{is} * TEST_{is} + \beta_4 X_{is} + \beta_{sg} + \varepsilon_{isg} \qquad (4)$$

Here, $Coll_{isg}$ is a dummy that equals one if student i who entered the STAR experiment in school s and grade g ever attended college. $TEST_{is}$ is the average of student i's non-missing kindergarten through third grade math and reading test scores, normalized to mean zero and standard deviation of one. Results are in Table 6 (Column 2).

First looking to the coefficient on test scores, in STAR a one-standard deviation

[15] We regress an indicator for college attendance against the average scores in multiple standardized tests administered when the subjects were between ages six and nine. Scores are normalized (within age) to mean zero and standard deviation one. We measure college attendance by 2006, when the children were 25 to 29 years old.

increase in K-3 scores is associated with a 17 percentage-point increase in the probability of attending college.[16] This is very similar to the relationship estimated among the children of the NLSY. The estimated coefficient on the interaction term between small class assignment and average test score is zero, indicating that scores have no differential predictive power for postsecondary attendance across students in small and regular classes. Similarly, the estimated coefficient on the small class indicator variable is also zero, suggesting that there is no additional boost to the likelihood a student attends postsecondary school from small class assignment after accounting for contemporaneous test scores (which are boosted by smaller classes). The pattern is similar if we replace college attendance with degree receipt (Columns 3-4). These findings indicate that short-term gains in cognitive test scores are indeed predictive of long-term benefits.

By contrast, we find that scores from tests administered after children left STAR are not nearly so predictive of its long-term effect. We estimate the equation just described, replacing contemporaneous scores with those obtained from tests administered in grades six through eight, three to five years after the experiment had ended. Now we see that, even after controlling for test scores, small-class assignment raises the likelihood of attending college by a statistically significant 2 percentage points. Further, the negative coefficient on the interaction term indicates that these subsequent test scores have less predictive power in small than regular classes. We conclude that scores recorded several years after the experiment do a significantly poorer job than contemporaneous scores in predicting the effect of the experiment on adult outcomes. One caveat to this analysis is that there could be omitted variables that are correlated both with assignment to a small class, test scores, and college attendance. If this were the case, then it might not be the contemporaneous test scores that are mediating the effect of small class assignment, but rather the omitted variables.

CONCLUSION

We estimate the effect of class size in early elementary school on postsecondary

[16] Results are unchanged if we exclude the school-by-wave fixed effects and demographics.

attainment. Assignment to a small class increases college attendance by 2.7 percentage points. Enrollment effects are largest among black students, students from low-income families, and high-poverty schools, indicating that class-size reductions during early childhood can help to close income and racial gaps in postsecondary attainment. Assignment to a small class also increases degree completion by 1.6 percentage points, with the effects concentrated in high-earning fields such as business, economics, and STEM.

As a final check on the sensitivity of our main result to possible sources of bias, we conduct two exercises. First, we examine the extent that students missing name and date of birth could influence the results, given that the NSC uses these identifiers to match students to college enrollment data. We assign all students with a missing name or date of birth first as having enrolled in college and then as having not enrolled in college regardless of their observed enrollment status. After each of these imputations we re-estimate Equation (1). Imputing students with missing identifiers as enrolled (not enrolled) yields a point estimate of 0.017 (0.025) and standard error of 0.009 (0.011). These coefficients are somewhat attenuated relative to our main result of 0.027 (SE=0.011). However, this check shows that even if we impute the most extreme cases of possible bias due to missing identifiers, our result remains positive, statistically significant, and similar in magnitude to our main result.

Our final check is a back-of-the-envelope exercise to bound the possible upward bias that could be due to small class assignment inducing students out of colleges not participating in the NSC (e.g., for-profits) and into colleges that do participate (e.g., community colleges). Using the NSC participant list and IPEDS enrollment data, we calculate that 8.7 percent of first-time enrollment in Tennessee during 1998 is in for-profit colleges. If small classes induce all of these students out of for-profit institutions and into colleges that we observed in the NSC (an extreme assumption), then our estimated effect on college enrollment would be biased upward by 3.7 percentage points.[17] This upper bound on the upwards bias is larger than our observed treatment

[17] In other words, if we assume that none of the treatment group attends for-profit colleges but 8.7 percent of the control group does, the implied total college enrollment rate among the control group would be 0.422. This rate is 3.7 percentage points higher

effect. However, a somewhat more realistic estimate based on past studies of STAR would be to assume that the treatment induces 10 percent of students out of for-profit institutions and into colleges that we observe (Krueger and Whitmore, 2001). This would cause our estimates to be biased upwards by 0.4 percentage points. This excludes any possible attenuation bias due to classical measurement error in the unobserved nonprofit college attendance, and any possible downward bias due to small classes inducing non-college attenders into for-profit institutions. This is thus a source of potential upward bias that under a somewhat plausible worst case scenario would explain only a small fraction of our treatment effect.

Is the nearly three percentage-point increase due to reduced class size that we estimate a large effect? To put this effect in context, we compare the estimate to those of other interventions that boost postsecondary attainment. We focus on the results of randomized trials when possible, turning to plausibly-identified quasi-experiments where no controlled experiment has been conducted. Deming and Dynarski (2010) provide a review of this literature, from which much of this information is drawn. We focus on evaluations of discrete, replicable interventions. We deliberately ignore several excellent papers that demonstrate that schools or teachers "matter" for postsecondary attainment, since they do not identify the effect of a manipulable parameter of the education production function (e.g., Deming et al., 2011, Chetty et al., 2011).

Two small experiments have tested the effect of intensive preschool on long-term outcomes. Abecedarian produced a 22 percentage-point increase in the share of children who eventually attended college. The Perry Preschool Program had no statistically significant effect on postsecondary outcomes (Anderson, 2008). The subjects in these experiments were almost exclusively poor and black. Head Start, a less intensive preschool program, increases college attendance by 6 percentage points (Deming, 2009), with larger effects for blacks and females (14 and 9 percentage points, respectively). Upward Bound provides at-risk high-school students with increased instruction, tutoring and counseling. The program had no detectable effect on the full

---

than the observed attendance rate among the control group (excluding for-profit colleges) of 0.385.

sample of treated students, but it did increase college attendance among students with low educational aspirations by 6 percentage points (Seftor, Mamun, & Schirm, 2009).

There are no experimental estimates of the effect of financial aid on college entry. However, there are several well identified quasi-experimental studies showing that student aid can boost postsecondary enrollment by several percentage points depending on how much aid is provided (Deming & Dynarski, 2010). Another way of increasing college enrollment is by assisting students with the administrative requirements of enrolling in college. Bettinger et al. (2012) randomly assign families to a low-cost treatment that consists of helping them to complete the FAFSA, the lengthy and complicated form required to obtain financial aid for college. Their intervention increases enrollment by eight percentage points.

The costs of the above interventions vary dramatically. We create an index of cost effectiveness for increasing college enrollment by dividing each program's costs by the proportion of treated students it induces into college.[18] Head Start costs $8,000 per child. Given the 6 percentage-point effect noted above, the amount spent by Head Start to induce a single child into college is therefore $133,333 (=$8,000/0.06). For Abecedarian, the figure is $410,000 (=$90,000/0.22). The cost of reduced class size is $12,000 per student, larger than that of Head Start but considerably smaller than that of Abcedarian. The amount spent in STAR to induce a single child into college is $400,000 (=$12,000/0.03). If the program could be focused on students in the poorest third of schools (the subpopulation that most closely matches that of the preschool interventions) then the cost drops to $171,000 per student induced into college.

Upward Bound costs $5,620 per student. If the program could be targeted to students with low educational aspirations, the implied cost of inducing a single student into college is $93,667 (=$5,620/0.06). Dynarski (2003) examines the effect of the elimination of the Social Security Student Benefit Program, which paid college scholarships to the dependents of deceased, disabled and retired Social Security

[18] All costs in this section are in 2007 dollars and come from Deming and Dynarski (2010) unless otherwise indicated. The costs for the early childhood programs and STAR have been discounted back to age zero using a 3 percent discount rate. Costs of the high school and college interventions have not been discounted.

beneficiaries. Eligible students were disproportionately black and low-income. The estimates from that paper indicate that about two-thirds of the treated students who attended college were inframarginal, while the other third was induced into the college by the $7,000 scholarship. These estimates imply that three students are paid a scholarship in order to induce one into college. The cost per student induced into college is therefore $21,000. Finally, the cost per treated subject in the FAFSA experiment (Bettinger et al., 2012) was $88 for an implied cost per student induced into college of $1,100 (=$88/0.08).

A fair conclusion from this analysis is that the effects we find in this paper of class size on college enrollment alone are not particularly large given the costs of the program. If focused on students in the poorest third of schools, then the cost-effectiveness of class size reduction is within the range of other interventions. There is no systematic evidence that early interventions pay off more than later ones when the outcome is limited to increased college attendance.

In addition to estimating the effects of reduced class size during childhood on educational attainment, the results in our paper shed light on the relationship between the short-and long-term effects of an educational intervention. We find that the short-term effect of small class assignment on test scores is an excellent predictor of its effect on adult educational attainment. In fact, under the assumption that there are no omitted variables correlated with small class assignment, test scores, and college enrollment, the effect of small classes on college attendance is completely "explained" by their positive effect on contemporaneous test scores. Further, the relationship between scores and postsecondary attainment is the same in small and regular classes; that is, the scores of children in the small classes are no less (or more) predictive of adult educational attainment than those of children in the regular classes. This is an important and policy-relevant finding, given the necessity to evaluate educational interventions based on contemporaneous outcomes.

A further contribution of this paper is to identify the effect of manipulating a single educational input on adult educational attainment. The early-childhood interventions for which researchers have identified lifetime effects (e.g., Head Start, Abecederian) are intensive and multi-pronged, including home visits, parental coaching and vaccinations. We cannot distinguish which dimensions of these treatments generate short-term effects on test scores, and whether they differ from the dimensions that generate long-term effects on adult wellbeing. By contrast, the effects we measure in this paper, both short-term and long-term, can be attributed to a well-defined and replicable intervention: reduced class size.

REFERENCES

Abdulkadiroglu, A., Angrist, J.D., Dynarski, S.M., Kane, T.J., & Pathak, P.A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. Quarterly Journal of Economics, 126 (2), 699–748.

Achilles, C.M. (1999). Let's put kids first, finally: Getting class size right. Thousand Oaks, CA: Corwin Press.

Anderson, M.L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association, 103 (484), 1481–1495.

Angrist, J.D. & Krueger, A.B. (1991). Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics, 106 (4), 979–1014.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., & Walters, C.R. (2012) Who benefits from KIPP? Journal of Policy Analysis and Management, 31 (4), 837–860.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of Econometrics, 121, 343–375.

Bailey, M.J. & Dynarski, S.M. (2011). Gains and gaps: A historical perspective on inequality in college entry and completion. In G. Duncan & R. Murnane (Eds.), Wither opportunity: Rising inequality, schools, and children's life chances. New York: Russell Sage.

Barrow, L., Richburg-Hayes, L., Rouse, C.E., & Brock, T. (2009). Paying for performance: The education impacts of a community college scholarship program for low-income adults. Working Paper No. 2009-13. Chicago: Federal Reserve Bank of Chicago.

Belley, P. & Lochner, L. (2007). The changing role of family income and ability in determining educational achievement. Journal of Human Capital, 1 (1), 37–89.

Bettinger, E.P., Long, B.T., Oreopoulos, P. & Sanbonmatsu, L. (2012) The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. Quarterly Journal of Economics, 127 (3), 1205–1242.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In E. Leamer & J.J. Heckman (Eds.) Handbook of Econometrics. XXcity: etc.

Bound, J., Lovenheim, M. & Turner, S.E. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. American Economic Journal: Applied Economics, 2 (3), 1–31.

Bowen, W.G., Chingos, M.M., & McPherson, M.S. (2009). Crossing the finish line: Completing college at America's public universities. Princeton, N.J.: Princeton University Press.

Cascio, E.U. & Staiger, D. (2012). Knowledge, test, and fadeout in educational interventions. NBER Working Paper No 18038. Cambridge, MA: National Bureau of Economic Research.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W. & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence From Project Star. Quarterly Journal of Economics, 126 (4), 1593–1660.

Dee, T.S. (2004). Are there civic returns to education? Journal of Public Economics, 88, 1697–1720.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics, 1 (3), 111–

134.

Deming, D. & Dynarski, S.M. (2010). Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. In P. Levine & D. Zimmerman (Eds.), Targeting investments in children: Fighting poverty when resources are limited. Chicago: University of Chicago Press.

Deming, D., Hastings, J., Kane, T., & Staiger, D. (2011). School choice, school quality and postsecondary attainment. NBER Working Paper No 17438. Cambridge, MA: National Bureau of Economic Research.

Dobbie, W. & Fryer, R.G. (2011). Are high quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. American Economic Journal: Applied Economics, 3 (3), 158–187.

Dynarski, S.M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. American Economic Review, 93 (1), 279–288.

Dynarski, S.M., Hemelt, S.W. & Hyman, J.M. (2012). Data watch: Using National Student Clearinghouse data to track postsecondary outcomes. Working Paper, University of Michigan.

Finn, J.D. & Achilles, C.M. (1990). Answers and questions about class size: A statewide experiment. American Educational Research Journal, 27, 557–577.

Folger, J. & Breda, C. (1989). Evidence from Project STAR about class size and student achievement. Peabody Journal of Education, 67, 17–33.

Fredriksson, P., Ockert, B., & Oosterbeek, H. (2013). Long-term effects of class size. Quarterly Journal of Economics, 128 (1), 249–285.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. American Economic Review, 92 (4), 999–1012.

Hamermesh, D.S. & Donald, S.G. (2008). The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. Journal of Econometrics, 144, 479–491.

Hoxby, C.M. & Murarka, S. (2009). Charter schools in New York City: Who enrolls and how they affect student achievement. NBER Working Paper No 14852. Cambridge, MA: National Bureau of Economic Research.

Krueger, A.B. (1999). Experimental estimates of education production functions. Quarterly Journal of Economics, 114, 497–532.

Krueger, A.B. & Whitmore, D.M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. Economic Journal, 111, 1–28.

Krueger, A.B. & Whitmore, D.M. (2002). Would smaller classes help close the black-white achievement gap? In J.E. Chubb & T. Loveless (Eds.) Bridging the Achievement Gap. Washington: Brookings Institution Press.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. Review of Economic Studies, 72, 189–221.

Ludwig, J. & Miller, D.L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics, 122 (1), 159–208.

National Center For Education Statistics (2010). Integrated Postsecondary Education Data System (IPEDS). Washington, D.C.: U.S. Department of Education.

National Science Foundation (2011). Science and Engineering Degrees: 1966-2008. Detailed Statistical Tables NSF 11-316. Arlington, VA: National Center for Science and Engineering Statistic.

Roderick, M., Nagaoka, J. & Allensworth, E. (2006). From high school to the future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from 4-year colleges. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Ruggles, S., Alexander, J.T., Genadek, K., Goeken, R., Schroeder, MB., & Sobek, M. (2010). Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database], Minneapolis: University of Minnesota.

Schanzenbach, D.W. (2006). What have researchers learned from Project STAR? Brookings Papers on Education Policy, 2006(1), 205-228.

Schweinhart, L.J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool study through age 40. Ypsilanti, MI: High/Scope Press.

Seftor, N.S., Mamun, A. & Schirm, A. (2009). The impacts of regular upward bound on postsecondary outcomes 7-9 years after scheduled high school graduation: Final report. Princeton, NJ: Mathematica Policy Research.

Word, E., Johnston, J., Bain, H.P., Fulton, B.D., Zaharias, J.B., Achilles, C.M., Lintz, M.N., Folger, J. & Breda, C. (1990). The state of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985-1990. Nashville: Tennessee State Department of Education.

Table 1. Means of Demographics and Outcome Variables by Class Size

	Regular Class	Small Class	Regression Adjusted Difference	
	(1)	(2)	(3)	
Demographics				
White	0.620	0.660	-0.003	(0.005)
Female	0.471	0.473	-0.000	(0.011)
Free Lunch	0.557	0.521	-0.015	(0.011)
College attendance				
Ever attend	0.385	0.420	0.027	(0.011)
Ever attend full-time	0.278	0.300	0.013	(0.011)
Enrolled On-Time	0.274	0.308	0.024	(0.011)
Number of Semesters				
Attempted	3.07	3.39	0.219	(0.133)
Attempted, conditional on attending	7.98	8.08	0.132	(0.209)
Degree Receipt				
Any degree	0.151	0.174	0.016	(0.009)
Associates	0.027	0.034	0.007	(0.004)
Bachelors or higher	0.124	0.141	0.009	(0.008)
Degree Type				
STEM, business or economics field	0.044	0.060	0.013	(0.006)
All other fields	0.085	0.094	0.003	(0.006)
First Attended				
2-year	0.215	0.245	0.025	(0.009)
Public 4-year	0.127	0.132	0.005	(0.007)
Private 4-year	0.042	0.043	-0.003	(0.004)
Number of Schools	79			
Number of Students	8,316	2,953		

Notes: Column (3) controls for school-by-wave fixed effects and demographics. Standard errors, in parentheses, are clustered by school.

Table 2. The Effect of Class Size on College Attendance - Linear Probability Models

	Total		White	Black	No Free Lunch	Free Lunch	Tercile of Poverty Share				P-value: High vs. Middle/Low
							High	Middle	Low	Middle & Low	
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
College Attendance											
Ever attend	0.028	0.027	0.011	0.058	0.010	0.044	0.073	-0.010	0.022	0.006	0.008
	(0.012)	(0.011)	(0.013)	(0.022)	(0.017)	(0.015)	(0.021)	(0.017)	(0.018)	(0.012)	
	0.385		*0.432*	*0.308*	*0.563*	*0.272*	*0.262*	*0.417*	*0.476*	*0.446*	
Ever attend full-time	0.014	0.013	-0.000	0.037	0.000	0.025	0.048	-0.012	0.008	-0.003	0.048
	(0.011)	(0.011)	(0.013)	(0.021)	(0.016)	(0.014)	(0.022)	(0.015)	(0.018)	(0.012)	
	0.278		*0.317*	*0.212*	*0.440*	*0.175*	*0.173*	*0.297*	*0.363*	*0.330*	
Enrolled On-Time	0.025	0.024	0.018	0.036	0.025	0.024	0.047	0.007	0.023	0.015	0.228
	(0.012)	(0.011)	(0.013)	(0.021)	(0.017)	(0.014)	(0.023)	(0.017)	(0.018)	(0.013)	
	0.274		*0.321*	*0.197*	*0.449*	*0.163*	*0.163*	*0.296*	*0.363*	*0.329*	
Demographics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Number of Schools	79		79		79		24	29	26	55	
Number of Students	11,269	11,269	7,160	4,109	4,454	6,815	3,681	3,784	3,804	7,588	

Notes: All regressions control for school-by-entry-wave fixed effects. Demographics include race, sex and free lunch status. Standard errors, in parentheses, are clustered by school. Control means are in italics below standard errors.

Table 3. The Effect of Class Size on College Choice - Linear Probability Models

		Tercile of Poverty Share		P-value: High vs. Middle/Low
	Total	High	Middle & Low	
Dependent variable	(1)	(2)	(3)	(4)
College attendance	0.027	0.073	0.006	0.008
	(0.011)	(0.021)	(0.012)	
	0.385	*0.262*	*0.446*	
First Attended:				
2-year	0.025	0.063	0.007	0.009
	(0.009)	(0.019)	(0.010)	
	0.215	*0.162*	*0.242*	
Public 4-year	0.005	0.009	0.003	0.690
	(0.007)	(0.011)	(0.010)	
	0.127	*0.070*	*0.156*	
Private 4-year	-0.003	0.001	-0.004	0.491
	(0.004)	(0.004)	(0.005)	
	0.042	*0.030*	*0.049*	
Ever Attended:				
Out of state	0.013	0.029	0.006	0.197
	(0.009)	(0.013)	(0.012)	
	0.138	*0.100*	*0.157*	
Selective	0.009	0.007	0.011	0.839
	(0.009)	(0.016)	(0.011)	
	0.184	*0.090*	*0.231*	
Number of Schools	79	24	55	
Number of Students	11,269	3,681	7,588	

Notes: All regressions control for school-by-entry-wave fixed effects and demographics including race, sex, and free lunch status. Standard errors, in parentheses, are clustered by school. Control means are in italics below standard errors.

Table 4. The Effect of Class Size on Persistence and Degree Receipt - Linear Probability Models

		Tercile of Poverty Share		P-value: High vs. Middle/Low
	Total	High	Middle & Low	
Dependent variable	(1)	(2)	(3)	(4)
Number of Semesters Attempted	0.22	0.32	0.19	0.651
	(0.13)	(0.26)	(0.15)	
	3.07	*1.91*	*3.65*	
Receive Any Degree	0.016	0.011	0.019	0.624
	(0.009)	(0.012)	(0.012)	
	0.151	*0.071*	*0.191*	
Highest Degree				
Associates	0.007	0.007	0.007	0.918
	(0.004)	(0.006)	(0.006)	
	0.027	*0.013*	*0.034*	
Bachelors or higher	0.009	0.003	0.012	0.532
	(0.008)	(0.011)	(0.010)	
	0.124	*0.058*	*0.157*	
Degree Type				
STEM field	0.005	0.000	0.008	0.194
	(0.003)	(0.004)	(0.004)	
	0.019	*0.008*	*0.024*	
Business or economics field	0.007	0.001	0.011	0.189
	(0.005)	(0.004)	(0.006)	
	0.026	*0.012*	*0.033*	
All other fields	0.003	0.013	-0.000	0.279
	(0.006)	(0.008)	(0.008)	
	0.085	*0.039*	*0.108*	
STEM, business or economics field	0.013	0.001	0.019	0.092
	(0.006)	(0.006)	(0.008)	
	0.044	*0.020*	*0.057*	
Number of Schools	79	24	55	
Number of Students	11,269	3,681	7,588	

Notes: All regressions control for school-by-entry-wave fixed effects and demographics including race, sex, and free lunch status. Standard errors, in parentheses, are clustered by school. Control means are in italics below standard errors.

Table 5. Examining Whether Heterogeneity is in Treatment Effects or Dosage

	First Stage	Reduced Form	Two-Stage-Least-Squares	Control Mean
	(1)	(2)	(3)	(4)
Everyone	0.643	0.006	0.009	0.385
(n=11,269)	(0.016)	(0.003)	(0.005)	
High Poverty Share	0.602	0.017	0.028	0.262
(n=3,681)	(0.025)	(0.006)	(0.010)	
Middle/Low Poverty Share	0.662	0.001	0.002	0.446
(n=7,588)	(0.019)	(0.004)	(0.005)	
Black	0.589	0.014	0.024	0.308
(n=4,109)	(0.019)	(0.006)	(0.010)	
White	0.669	0.003	0.004	0.432
(n=7,160)	(0.019)	(0.004)	(0.006)	
Free Lunch	0.628	0.010	0.016	0.272
(n=6,815)	(0.015)	(0.004)	(0.007)	
Non-Free Lunch	0.665	0.002	0.003	0.563
(n=4,454)	(0.024)	(0.005)	(0.008)	

Notes: All regressions control for school-by-entry-wave fixed effects and demographics including race, sex, and free lunch status. Standard errors, in parentheses, are clustered by school.

Table 6. Examining Whether Short-Term Gains Predict Long-Term Gains - Linear Probability Models

	College Enrollment		Degree Receipt	
	(1)	(2)	(3)	(4)
Mean K-3 Test Score				
Small class	0.027	0.002	0.016	0.001
	(0.011)	(0.009)	(0.009)	(0.009)
Test score		0.169		0.096
		(0.006)		(0.007)
Small class * test score		-0.008		0.000
		(0.010)		(0.008)
Mean 6-8 Test Score				
Small class	0.027	0.020	0.016	0.010
	(0.011)	(0.010)	(0.009)	(0.008)
Test score		0.230		0.141
		(0.005)		(0.006)
Small class * test score		-0.014		0.009
		(0.008)		(0.008)
Control Mean	0.385	0.385	0.151	0.151
Number of Students	11,269	11,269	11,269	11,269

Notes: All regressions control for school-by-entry-wave fixed effects and demographics including race, sex, and free lunch status. Missing test-score indicators included for students with no test scores in grade range. Standard errors, in parentheses, are clustered by school.

Appendix Table 1. Student Demographics by School Poverty Share

	High Poverty	Middle Poverty	Low Poverty	Middle/Low Poverty
	(1)	(2)	(3)	(4)
White	0.253	0.746	0.881	0.814
Female	0.471	0.475	0.469	0.472
Free Lunch	0.855	0.504	0.292	0.398
Number of Schools	24	29	26	55
Number of Students	3,681	3,784	3,804	7,588

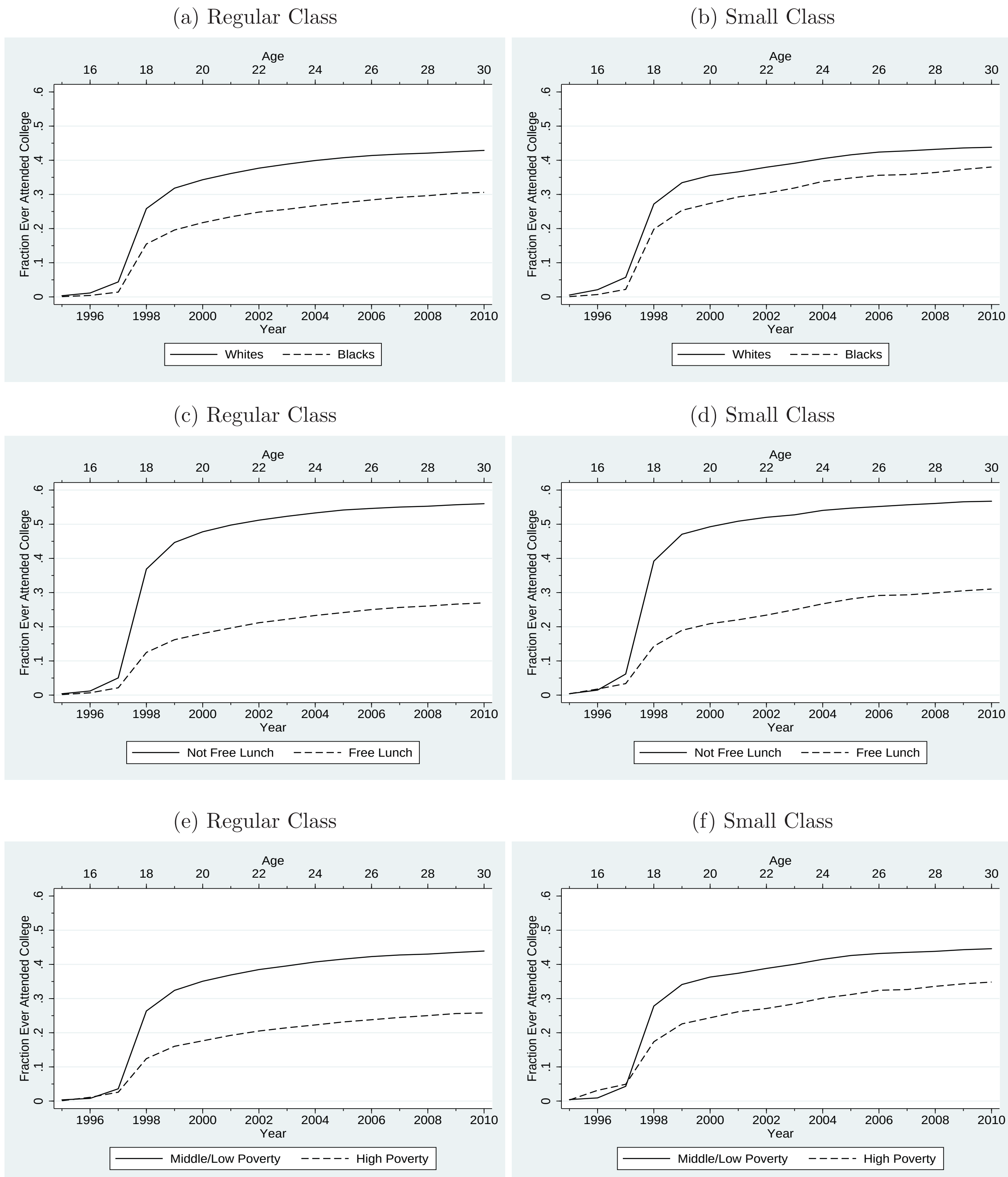
Notes: School poverty share is measured as the fraction of the school that is eligible for a subsidized lunch.

Appendix Table 2. The Effect of Class Size Censoring to Match IRS Data Span - Linear Probability Models

	Baseline - All Years of Enrollment	Exclude Pre-1999 Enrollment	Exclude Post-2007 Enrollment	Include 1999-2007 Enrollment Only
Dependent variable	(1)	(2)	(3)	(4)
Ever attend	0.027	0.018	0.023	0.015
	(0.011)	(0.011)	(0.011)	(0.011)
	0.385	*0.369*	*0.372*	*0.357*
Number of Students	11,269	11,269	11,269	11,269

Notes: All regressions control for school-by-entry-wave fixed effects and demographics including race, sex, and free lunch status. Standard errors, in parentheses, are clustered by school. Control means are in italics below standard errors.

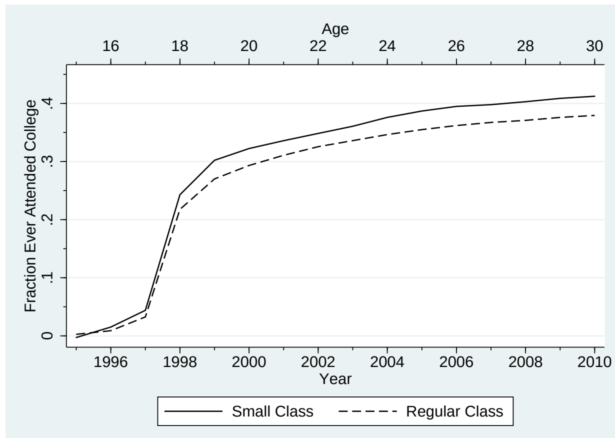
Figure I: The Effect of Class Size on Racial and Income Gaps in Postsecondary Attainment



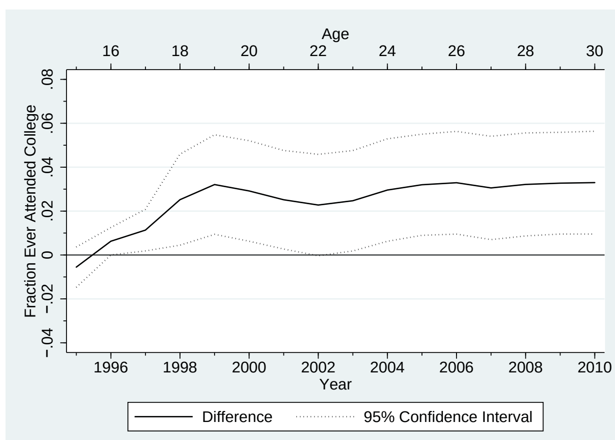
Notes: Figures (a), (c), and (e) plot the fraction ever attended college by year for STAR students assigned to regular size classes, and figures (b), (d), and (f) for STAR students assigned to small classes. Figures (a) and (b) compare college attendance by race, figures (c) and (d) by free lunch status, and figures (e) and (f) by school poverty share.

Figure II: College Attendance Over Time, By Class Size

(a) Fraction Ever Attended College



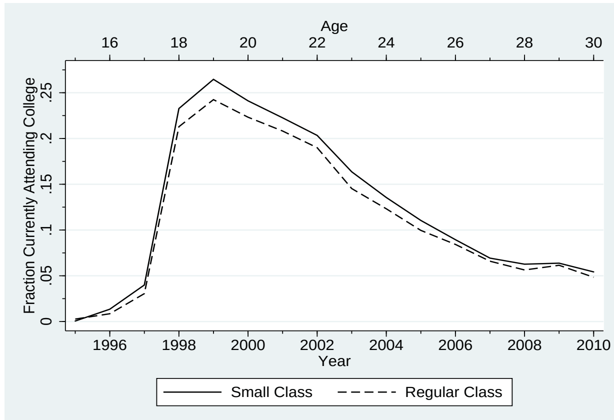
(b) Difference Between Small and Regular



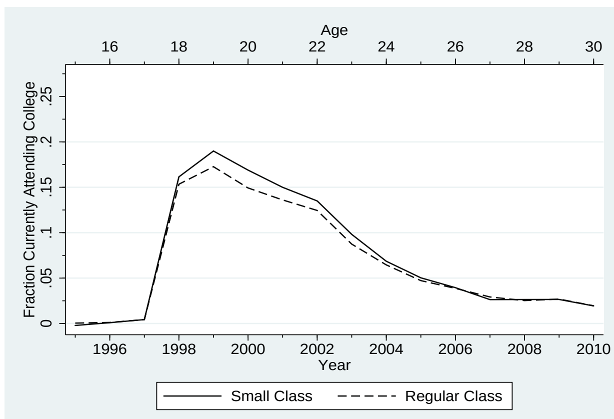
Notes: Figure (a) plots the mean fraction ever attended college by year for students assigned to small vs. regular size classes. It controls for both school-by-wave fixed effects and demographics, including race, sex and free lunch status. Figure (b) plots the difference and its 95% confidence interval by year. Standard errors are clustered by school.

Figure III: Fraction Currently Enrolled in College Over Time, By Class Size and Enrollment Status

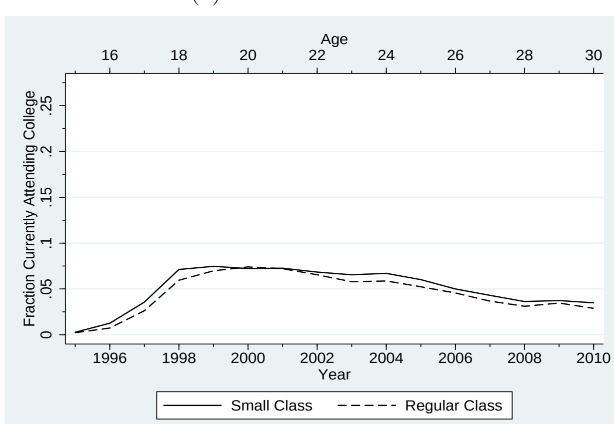
(a) Any Enrollment Status



(b) Full-Time Status



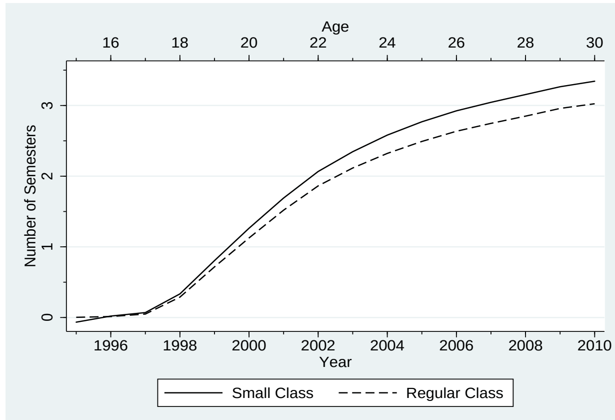
(c) Part-Time Status



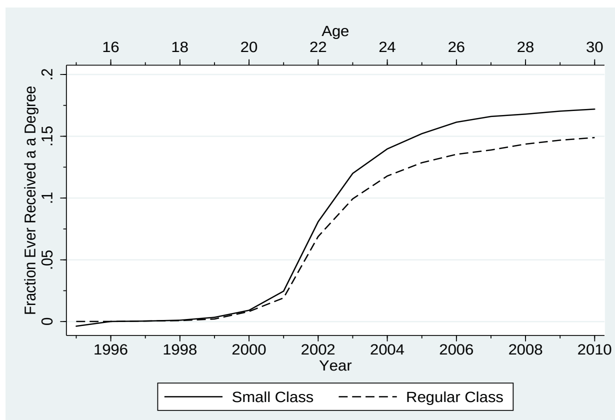
Notes: Figures plot the fraction currently attending college by year for STAR students assigned to small vs. regular size classes. All figures control for both school-by-wave fixed effects and demographics, including race, sex and free lunch status.

Figure IV: Postsecondary Persistence and Degree Receipt Over Time, By Class Size

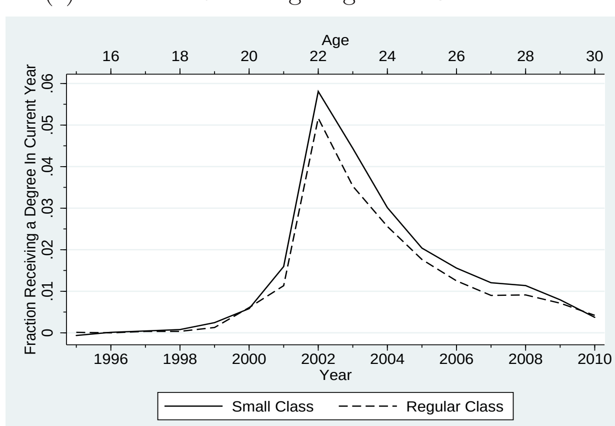
(a) Cumulative Number of Semesters Attended



(b) Fraction Ever Received A Degree



(c) Fraction Receiving Degree in Current Year



Notes: Figure (a) plots the mean cumulative number of semesters attended by year for STAR students assigned to small vs. regular size classes. Figure (b) plots the mean fraction ever receiving any postsecondary degree (associates or higher). Figure (c) plots the mean fraction receiving any postsecondary degree in the current year. All figures control for both school-by-wave fixed effects and demographics, including race, sex and free lunch status.