NBER WORKING PAPER SERIES

DO COLLEGE-PREP PROGRAMS IMPROVE LONG-TERM OUTCOMES?

C. Kirabo Jackson

Working Paper 17859
http://www.nber.org/papers/w17859

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2012

This paper includes material circulated in, and replaces the unpublished NBER working paper #15722 titled "The Effects of an Incentive-Based High-School Intervention on College Outcomes." The author is grateful for helpful comments and suggestions from David Figlio and Karl Scholz. All errors are my own. This research was made possible through data provided by the University of Texas at Dallas Education Research Center. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Do College-Prep Programs Improve Long-Term Outcomes?
C. Kirabo Jackson
NBER Working Paper No. 17859
February 2012
JEL No. H0,I20,J01

ABSTRACT

I analyze the longer-run effects of a college-preparatory program implemented in inner-city schools that included payments to eleventh- and twelfth- grade students and their teachers for passing scores on Advanced Placement exams. Affected students attended college in greater numbers, were more likely to remain in college beyond their first year, more likely to earn a college degree, more likely to be employed, and earned higher wages. This is the first credible evidence that implementing college-preparatory programs in existing urban schools can improve both the long-run educational and labor market outcomes of disadvantaged students.

C. Kirabo Jackson
Northwestern University
School of Education and Social Policy
2040 Sheridan Road
Evanston, IL 60208
and NBER
kirabo-jackson@northwestern.edu

Do College-Prep Programs Improve Long-Term Outcomes?[1]

C. Kirabo Jackson
Northwestern University, IPR, and NBER
kirabo-jackson@northwestern.edu

I analyze the longer-run effects of a college-preparatory program implemented in inner-city schools that included payments to eleventh- and twelfth- grade students and their teachers for passing scores on Advanced Placement exams. Affected students attended college in greater numbers, were more likely to remain in college beyond their first year, more likely to earn a college degree, more likely to be employed, and earned higher wages. This is the first credible evidence that implementing college-preparatory programs in existing urban schools can improve both the long-run educational and labor market outcomes of disadvantaged students.

The economic returns to education have been rising in the United States. In 1979, college-educated adults earned seventy-five percent more than high school graduates, while by 2003, college-educated adults earned two-hundred and thirty percent more than high school graduates (Rouse and Barrow 2006). At the same time, college enrollment is much lower among ethnic-minority high school graduates and among students from low-income families.[2] As such, these increasing returns have important implications for earnings differences across groups and for intergenerational income mobility in the United States (Turner 2004; Acemoglu and Autor 2010). Motivated by these facts, federal and state governments make significant expenditures on programs aimed at increasing college preparation among disadvantaged populations.[3] Because students often self-select into such programs, rigorous evaluations are scarce and there are none investigating the effects of college prep-programs on long-run outcomes. This paper aims to fill this void in the literature by analyzing the long-run educational and labor market effects of the Advanced Placement Incentive Program (APIP). The APIP is a high school intervention that includes cash incentives for both teachers and students for passing scores earned on AP exams,

[1] This paper includes material circulated in, and replaces the unpublished NBER working paper #15722 titled "The Effects of an Incentive-Based High-School Intervention on College Outcomes." The author is grateful for helpful comments and suggestions from David Figlio and Karl Scholz. All errors are my own. This research was made possible through data provided by the University of Texas at Dallas Education Research Center. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.

[2] Approximately 70 percent of white high school graduates or GED holders between the ages of 25 and 29 enrolled in some college program. The corresponding figures are roughly 60 and 50 percent for blacks and Hispanics, respectively (Current Population Survey 2010). Roughly eighty percent of the high school graduates from the top income quartile attend some college compared to under sixty percent of those from the lowest income quartile (Ellwood and Kane 2000).

[3] In 2011, the federal government spent over $650 million to promote college-preparation programs for low-income students such as Upward Bound and Gear UP (U.S. Department of Education). Additionally, several states have pushed to expand Advanced Placement and International Baccalaureate programs (Lerner & Brand, 2008).

teacher training, curricular oversight, and test-prep sessions. To my knowledge, this analysis of the APIP represents the first rigorous investigation into whether college preparation programs confer long-run educational and labor market benefits.[4]

There are few existing studies on the effects of college-prep programs that use experimental or quasi-experimental research designs, and the results are mixed. Seftor, Mamun, and Schirm (2009) conducted a randomized evaluation of Upward Bound, an intensive college-prep program for ninth- through twelfth-grade students that includes additional instruction, tutoring, and counseling during the school year, and intensive instruction during the summer. The authors found no overall effect on postsecondary enrollment or completion. However, Cahlan (2008) raises serious questions about these findings because the comparison group may have also received college-prep programming, there was attrition bias, and the sample was weighted heavily to one school with improper implementation. Another study is Jackson (2010) that examined the short-run effects of the APIP and found that the program improves SAT performance and college matriculation. I am aware of no other rigorous studies on college-prep programs, and none that look at long-run labor market outcomes.[5]

The APIP, like most college-preparatory programs such as Upward Bound and GEAR UP, includes academic instruction, tutoring, and counseling. What distinguishes the APIP from other programs is that it *also* uses cash incentives for both students and teachers to induce greater participation in AP programs, increase teacher encouragement for students to take AP courses, and increase student and teacher effort in AP courses and exams. On the student side, theory predicts these cash incentives should lead to increased AP participation and effort, if students are myopic, reduce time studying because of the need to work, or face social pressures not to study. On the teacher side, cash incentives should increase AP participation and instruction quality if teachers reduce class sizes, or reduce time spent working with AP students on other activities. As such, the use of cash rewards may make this program particularly likely to succeed.[6]

---

[4]Recent findings indicate that students, who disproportionately attend urban schools, have improved educational outcomes when given the opportunity to attend better schools (Deming et al. 2011; Abdulkadiroğlu et al. 2011; Booker et al. 2011; Dobbie & Fryer 2011; Jackson 2010). However, it has proven difficult to reduce educational disparities by improving the existing schools to which most disadvantaged students are consigned (Murnane 2008).

[5] Kemple and Willner (2008) conduct a randomized evaluation of career academies, which focus on vocational skills through work internships and career-related experiences in addition to academics. They found no effect on educational attainment, but positive effects on labor market outcomes. Because career academies are oriented toward career-relevant skills, these important findings do not speak to the effects of college-preparation programs.

[6] For incentive-based interventions, such as the APIP, some psychologists argue that external rewards can supplant intrinsic motivation, such that performance may be worse after incentives are removed than if they had never been

Using K-12 education data linked to college records (both in Texas and outside of Texas) and unemployment insurance records in Texas in 2010, I investigate how the APIP, which was administered to high school juniors and seniors, affected their (1) college attendance, (2) sophomore-year college persistence, (3) college completion, (4) labor market participation, and (5) labor market earnings. Because the APIP was not implemented in all interested high schools at once, there is variation in the timing of APIP adoption within the sample of interested schools. This allows for a quasi-experimental difference-in-difference strategy— comparing the change in outcomes between observationally similar students from the same high school before and after APIP adoption to the change in outcomes across cohorts from other high schools that did not adopt the APIP over the same time period. Comparing cohorts from the same school removes differences in unobserved attributes that make one student take AP courses while another does not and also differences in unobserved attributes that make students from certain schools more likely to be successful than others. Using changes at similar schools over the same period as a comparison removes the effects of state policies and influences that may coincide with adoption at some schools. Because the timing of adoption was nonrandom, I present several tests to show that the results are not driven by (a) selective migration to treated schools, (b) preexisting trends, (c) changes in other school inputs, or (d) changes in school leadership.

I find that affected students took and passed more AP course and exams, and enrolled in college in greater numbers. Most of this increase occurred at four-year colleges and private universities. Affected students were also more likely to persist in college, to earn more college credits, and slightly more likely to earn a bachelor's degree. In addition, affected students were more likely to be employed and earned higher wages. Conservative estimates suggest a benefit-cost ratio greater than 10. Consistent with a causal relationship, (a) improvements occur *after* adoption rather than before, (b) effects are strongest for students most likely to be affected, (c) students at treated schools who were unlikely to be treated experienced no effects, (d) schools with higher treatment intensity experienced larger effects, and (e) outcomes that should not have been affected by the APIP were unrelated to adoption.

---

introduced. This notion is discussed in (Deci and Ryan 1985) and has been popularized in (Kohn 1999). For a balanced meta-analysis of this literature, see (Cameron and Pierce 1994). Based on actual field studies (as opposed to lab studies), Angrist and Lavy (2009) find that student incentives improve outcomes for girls, and Lavy (2009) and Figlio and Kenny (2007) find that teacher incentives are associated with contemporaneous improvements in achievement for all students. According to Angrist, Lang, and Oreopoulos (2009), cash rewards in conjunction with additional supports for academic achievement among college students lead to higher GPAs for female students.

These effects conceal important heterogeneity. Among students likely to take AP courses, there was a 1.7-percentage point increase in graduating from college, a 2.7-percentage point increase in the likelihood of being employed in 2010, and a 5.7-percent increase in earnings conditional on employment. The results also reveal sizable benefits for Hispanic students, who experience about a 2.5-percentage point increase in college degree attainment and an 11-percent increase in earnings. The earnings increases for Hispanic and black students are large enough to reduce the black-white earnings gap by one third and to eliminate the Hispanic-white earnings gap entirely. The long-run effects were most significant in schools with established AP programs before APIP adoption and in schools with high-powered incentives; this suggests that increased supply of AP courses was not the driving mechanism, and that increased teacher and student effort associated with the incentives are important aspects of the program's success.

These findings indicate that college-prep programs that both maintain high standards and increase participation in rigorous courses can improve college readiness and long-run educational and labor market outcomes. The results suggest that incentive-based programs that include resources to turn effort into achievement may have lasting positive effects, even after rewards are no longer provided. These findings also contribute to the debate on early versus late interventions (Cunha and Heckman 2007), because they show that an inexpensive program targeted to high school students is effective at increasing educational attainment and earnings. Finally, the results demonstrate that high-quality college-preparatory programs can improve the long-run economic well-being of disadvantaged students consigned to inner-city schools.

The remainder of this paper is structured as follows. Section II describes the APIP program, followed by the presentation of data in Section III. Then, Section IV discusses the program's empirical strategy. Section V presents the results, specification, and robustness tests, and finally Section VI concludes.

II. Description of the AP Incentive Program

AP courses are typically taken by students in eleventh and twelfth grades. The courses are intended to be "college level," and most colleges allow successful AP exam takers to use their scores to offset degree requirements. The AP program has 35 courses and examinations across 20 subject areas. The length of a course varies from one to two semesters. The cost per examination is $82, and a fee reduction of $22 is granted to those students with demonstrated financial need. AP exams are administered by the College Board, making teacher cheating unlikely. Exams are scored on a scale of through 5, where 3 and higher are regarded as passing grades. In Texas, AP courses are taught during regular class time and generally substitute for another course in the same subject, an elective course, or a free period. While AP courses count toward a student's high school GPA, they are beyond what is required for high school graduation and substitute for less demanding activities.[7]

The APIP is run by AP Strategies, a nonprofit organization based in Dallas, and is voluntary for schools, teachers, and students. The heart of the program is a set of financial incentives for teachers and students based on AP examination performance. The program also includes teacher training and curricular oversight conducted by the College Board and a curriculum that prepares students for AP courses in earlier grades. The APIP uses "vertical teams" of teachers. At the top of a vertical team is a lead teacher who instructs students and trains other AP teachers.[8] In addition to the AP courses taught at school, there may be extra time dedicated to AP training. For example, the APIP in Dallas includes special "prep sessions" for students once or twice a year, where up to 800 students gather at a single high school to take seminars from AP teachers as they prepare for their AP exams (Hudgins 2003).

The monetary incentives are intended to encourage participation and to induce effort in AP courses. AP teachers receive between $100 and $500 for each AP score of 3 or over earned by a high school junior or senior enrolled in their course and can receive discretionary bonuses of up to $1,000 based on results. In addition, lead teachers receive an annual salary bonus of $3,000 to $10,000, and a further $2,000 to $5,000 bonus opportunity based on results. While the amount paid per passing AP score and the salary supplements are well defined in each school, there is variation across schools in the amounts paid. Overall, the APIP can deliver a considerable increase in compensation for teachers.[9]

Students in eleventh and twelfth grades also receive incentives for performance. The program pays half of each student's examination fees, so that students on free or reduced lunch would pay $15 (instead of $30), while those who are not would pay $30 (instead of $60) per

[7] **Source:** Executive Vice-President of AP Strategies, and counselors at Dallas high schools.
[8] Vertical teams also include teachers whose grades precede those in which AP courses are offered. For example, a vertical team might create a seventh-grade math curriculum designed to prepare students for AP Calculus in twelfth grade. This may be important, given findings by (Jackson & Bruegmann, 2009) that teachers learn from their peers.
[9] One AP English teacher in Dallas had six students out of 11 score a 3 or higher on the AP examination in 1995, the year before the APIP was adopted. In 2003, when 49 of her 110 students received a 3 or higher, she earned $11,550 for participating in the program; this was a substantial increase in annual earnings (Mathews, 2004).

exam. Students receive between $100 and $500 for each score of 3 or above in an eligible subject for which they took the course. The amount paid per exam is well defined in each school, but there is variation across schools in the amount paid per passing AP exam. A student who passes several AP exams during junior and senior years can earn several hundred dollars. Because students must attend the AP courses *and* pass the AP exams to receive the rewards, the incentives are relatively difficult to game and are thus likely to increase student learning.

The total cost of the program ranges from $100,000 to $200,000 per school per year, depending on the size of the school and its students' propensity to take AP courses. The average cost per student in an AP class ranges from $100 to $300. Private donors pay for roughly 70 percent of the total costs, and the district covers the remainder. Districts pay for teacher training and corresponding travel, release time, and some of the supplies and equipment costs. Donors fund the cash rewards to students and teachers, stipends to teachers, bonuses to teachers and administrators for AP performance, and some of the supplies and equipment costs.[10]

As a rule, adoption of the APIP works as follows. First, interested schools approach AP Strategies and are placed on a list.[11] AP Strategies then tries to match interested schools to a donor. When a private donor approaches AP Strategies, he or she selects which schools to fund from within the list of interested schools. In most cases, the donor wants a specific district.[12] Once a willing group of schools has been accepted by the donor, preparations are made (such as training and creation of curricula) and the program is implemented the following calendar year.[13] Approximately two years are required to fully implement the APIP after a school expresses interest. Donors choose the subjects that are rewarded and ultimately determine the size of the financial rewards. While there are differences across schools, most schools reward English, mathematics, and the sciences.

There is variation in the timing of the introduction of the program across schools that I

[10] Today, districts can fund their contribution to the APIP using earmarked funds from the statewide AP incentive program and No Child Left Behind. However, in the first few years of the program, such funds were not available.
[11] There are a few exceptions. Schools in Austin were approached by a donor to adopt the APIP in 2007. In addition, five schools in Dallas secured a donor before approaching AP Strategies.
[12] For example, the first ten Dallas schools were chosen based on their proximity to AP Strategies; ST Microelectronics is located in the Carrolton-Farmers community and funded this district's schools; the Priddy Foundation specifically requested the Burkburnett and City View schools; anonymous donors specifically requested Amarillo and Pflugerville schools; the Dell Foundation (based in Austin) funds the Austin and Houston programs; and the remaining Dallas schools were funded by the O'Donnell Foundation to complete the funding of Dallas ISD.
[13] The seven schools that adopted the APIP in 2008, however, decided to have the pre-AP preparation portion of the program in place for at least a year before the rewards were provided.

exploit to identify the effect of the program. As illustrated in Figure 1, 58 schools adopted the APIP by 2008 (56 of which were early enough to have college outcomes). Because donor preferences determine the schools that will adopt the program in any given year, *among interested schools*, donor availability and preferences are the primary reasons for variation in the timing of program implementation.[14] To quote the Vice-President of AP Strategies, "Many districts are interested in the program but there are no donors. So there is always a shortage of donors." I argue that the exact timing of program adoption, within the group of willing schools, is orthogonal to *changes* in potentially confounding school characteristics. I test this assumption empirically in Section VI and show that it is likely valid.

III. The APIP Schools and the Data

To show how APIP schools differ from other schools in Texas, I present school-level summary statistics from the National Center for Education Statistics and the Texas Education Agency (TEA) in Table 1. Schools selected for the APIP were different from schools that had not been selected for the APIP. The APIP schools had average enrollments during 2000 through 2005 of 1836 students compared to 751 students for non APIP schools in Texas. During these years, 74 percent of the APIP schools were in large or mid-sized cities compared to under 20 percent for non-APIP schools. During these years, only 25 percent of students at APIP schools were white compared to 53 percent at non-APIP schools, and 10 percent of students had limited English proficiency at APIP schools compared to less than four percent at other schools. Both groups, however, have similar shares of economically disadvantaged students, reflecting the fact that Texas has both urban and rural poor.

The regression data contain individual student records from every public tertiary institution in Texas[15] between 1995 and 2010, every private institution in Texas between 2003 and 2010,[16] and most higher education institutions across all of the United States listed in the National Student Clearinghouse between 2008 and 2010 from the Texas Higher Education Coordinating Board. These data are linked with student-level high school and middle school data

[14] For example, in 2005, four high schools were chosen by The Michael and Susan Dell Foundation from a list of seven interested Houston schools. The remaining three schools may adopt the program at a later date.

[15] Texas has 145 institutions of higher learning. Of the public institutions, there are 35 universities, 50 community colleges, nine health-related institutions, four technical colleges, and three state colleges. On the private side, there are 39 universities, two junior colleges and three heath-related institutions.

[16] Because private school data are only available after 2003, I have run all models using only those cohorts that would have expected high school graduation after 2003, and the main results are largely the same.

(including standardized tenth-grade test scores that all students must take by state law) from the TEA for the years 1994 through 2007.[17] The standardized test scores are standardized to be mean zero with unit variance for each test administration. For each student, I use the most recent administration of the test (i.e., the year directly preceding expected exposure to the APIP).[18] These education data are linked to earnings and employment data for the fourth quarter in 2010 from the Texas Workforce Commission (TWC). These labor market data come from unemployment insurance records and provide information on earnings for employed Texas residents. Utilizing earnings data from only the fourth quarter in 2010 has a few benefits. First, because this time period was at the height of the Great Recession, any differences in skills that may manifest themselves in differences in labor market outcomes are likely to show up in these data. Second, using data across the same time period for all workers removes any potentially confounding effect associated with certain workers being employed at different points of the business cycle. Finally, using fourth-quarter earnings (as opposed to using annual earnings) allows one to observe workers who graduated in May of 2010 and were employed in the fourth quarter of that year ─ effectively increasing the sample size for this outcome by about twenty-five percent. The final dataset contains 2010 labor market outcomes, college outcomes, as well as high school and middle school data of all students who were in tenth grade between 1994 and 2007. Using the population of tenth graders allows me to account for attrition that may take place after APIP exposure in eleventh and twelfth grades.[19]

In Table 2, I present the pre- and post-APIP adoption summary statistics for schools that adopted the APIP by 2008 (*note: schools adopt the APIP at different times so the pre-adoption years differ across schools*). About 22.9 percent of students who were in tenth grade during the pre-adoption years took an AP course while in high school, compared to 30.4 percent in the post-adoption period. Similar increases in AP exam-taking were observed, where those in tenth grade during the pre period took 0.097 exams and those in the post period took 0.127 exams during high school. The tenth-grade math and reading scores were below zero for both periods, indicating that APIP schools were low-achieving schools on average. These scores were slightly lower after adoption than before, suggesting possible negative selection into APIP schools.

---

[17] TAKS (1994-2003) and TASP (2003-2007).

[18] The tenth-grade retention rate was approximately seven percent in Texas in 1995; among minorities, this figure was over 10 percent (http://www.tea.state.tx.us/reports/1996cmprpt/04retain.html).

[19] For example, if the APIP caused students to drop out of high school in eleventh grade, then using the population of juniors or senior would yield results that suffer from attrition bias.

The average student before adoption was in tenth grade in 1999 compared to 2003 for the post-adoption sample. As such, variables that increase with age, such as college attendance and earning, are difficult to compare without directly accounting for age (as is done in the regression analysis). However, to allow for a simple comparison, I compute enrollment by the time since expected high school graduation (not shown in the table). About 35, 46, and 53 percent of tenth graders in the pre-treatment cohorts were freshmen in college within one, two, and three years of expected high school graduation, respectively. The tenth graders in post-adoption cohorts were more likely to attend college, such that 41, 50, and 56 percent of tenth-grade students in post-treatment cohorts were freshmen in college within one, two, and three years of expected high school graduation, respectively.[20] The implied sophomore-year persistence (the share of students who were sophomores divided by the share who were freshmen) is 0.53 and 0.544 for the pre- and post-adoption cohorts, respectively.[21] Comparing figures for ever being a freshman to the enrollment figures above reveals that 59, 79, and 89 percent of college attendance occurs within one, two, and three years of expected high school graduation, respectively. Because analyzing college outcomes within four years of expected high school graduation ignores about one tenth of the variation in enrollment (and even more of the variation in college completion), I analyze college outcomes at any point in time and then control for cohort differences directly.

As the outcomes that change are a function of age, because age and treatment status are correlated, a simple treated versus un-treated comparison understates any effects of the APIP. In any case, among the untreated cohorts, 18 percent of high school sophomores eventually earn a college degree compared to about 14 percent of the treated (younger) cohorts. Looking to labor market outcomes, about 49 percent of tenth graders in the untreated cohorts were employed (according to unemployment insurance records) in 2010 and received mean three-month earnings of $8,029.60, while about 46 percent of tenth graders in the treated (younger) cohorts were in the labor market receiving mean three-month earnings of $4,895.20.

IV. Empirical Strategy

Before presenting the identification strategy, I discuss some methodological concerns

[20] Texas has the second-largest community college system in the US, so many students enroll in two-year colleges.
[21] The sophomore variables are not computed for the 2007 cohort because they would have been freshman in 2010 if they had enrolled directly after expected high school graduation. College graduation variables are computed for cohorts before 2004 because later cohorts are very unlikely to have graduated from a four-year college by 2010.

facing this and similar studies and then I present my proposed solutions. Because the APIP affects high school juniors and seniors, it influences the characteristics of students while still in high school, so that one must compare students who were similar *before* exposure to the APIP. As such, I compare the outcomes of students with similar attributes before exposure to APIP from the same high school, and I *do not* control for potentially endogenous covariates such as SAT scores or high school GPA.[22] The second methodological issue is how treatment is defined. Because students may enroll at APIP schools in eleventh and twelfth grades to benefit from the program, defining treatment based on actual school enrollment in these grades could be subject to selection bias. To avoid such bias, I use intention-to-treat (ITT) instead of whether a student is actually affected by the APIP. Specifically, I define ITT based on whether a student would be treated if she remained in her tenth-grade high school and were never held back a grade. For example, a student is intended for treatment if she is enrolled at a school in tenth grade in year t, and the school will have adopted the APIP by year $t+2$. Using ITT is advantageous because it is not endogenously determined by student selection into APIP schools in eleventh and twelfth grades, or subject to biases due to attrition or retention.[23] In addition, employing ITT yields a clean policy-relevant estimate of the effect of introducing the APIP to the target population.

IV.1 ***Identification Strategy***: The identification strategy is to compare the difference in outcomes across cohorts of students who attended the same high school before and after APIP adoption to the difference in outcomes between cohorts of students at schools that did not adopt the APIP over the same time period. Comparing students from the same high school addresses the concern that students at schools that adopt the APIP may differ from students who attend schools that do not adopt the APIP. By comparing cohorts, as opposed to students within cohorts, I address the concern that certain types of students tend to take AP courses and exams for unobserved reasons while others do not. Furthermore, by comparing the outcomes of students with the same tenth grade test scores and demographics, I address the concern that the incoming preparation of students may have changed in APIP schools after adoption of the program. Finally, this approach helps to account for potentially confounding statewide policies.[24]

---

[22] Many studies mistakenly control for endogenous variables to isolate the causal effect of AP exams.

[23] The downside of this measure is that it will not capture the full effect of the *treatment on the treated* since (1) students who leave APIP schools after tenth grade will not be treated but will be intended for treatment, (2) students who enter APIP schools after tenth grade will be treated but will not be intended for treatment, and (3) retained students, who should have graduated before APIP adoption, will be treated but will not be intended for treatment.

[24] For a description of such policies, see Appendix Note 2.

This strategy relies on the assumption that the difference in outcomes across cohorts for comparison schools is the same, in expectation, as the difference in outcomes across cohorts that adopting schools would have experienced if they had not adopted the APIP. For this to be plausible, the comparison schools must be similar to APIP-adopting schools. To ensure that this is the case, I restrict the estimation sample to those schools that adopted the APIP by 2008, using the change in outcomes for other APIP schools that did not yet have the opportunity to implement the program as the counterfactual change in outcomes. This sample restriction has two important benefits: (1) because APIP-willing schools are observationally similar, they likely share common time shocks; and (2) because APIP-willing schools are similarly motivated and interested, restricting the sample in this way avoids comparing schools with motivated principals to schools with unmotivated principals who have no interest in the program.[25]

Because I do not compare schools that adopt the APIP to those that do not, using a within-school estimation strategy controls for school selection on *unobserved* time-invariant characteristics, such as time-constant school enthusiasm or motivation. Identification relies on the assumption that the *timing* of APIP implementation is exogenous to other within-school *changes*. Since the timing of actual adoption relies on idiosyncratic donor preferences and availability, this assumption is plausible. However, because donor choices are not random, I cannot entirely rule out that the timing of adoption is uncorrelated with *changes* in school characteristics. As such, to assuage concern that *timing* of adoption may be endogenous, I identify those schools with which donors had prior relationships and verify that all the main results are robust to excluding these schools.[26] In addition, in Section V, I show that improvements only take place *after* APIP adoption; the timing of adoption is unrelated to other school changes or the timing of the arrival of a new principal; improvements are only experienced in AP-related outcomes and not other outcomes; and improvements are experienced almost entirely by students who were *ex-ante* likely to take AP courses and to be "treated." While none of these tests are dispositive individually, the weight of evidence indicates that the assumption of exogenous timing of adoption is probably valid.

This within-school cohort-based comparison is implemented by estimating the following

[25] Some schools adopted the APIP after 2007 and are therefore never treated in-sample for the purposes of analyzing college outcomes, but serve as comparison schools. The results are similar (albeit less precise) when using only schools that adopt the APIP in-sample, so the findings are not driven by the particular choice of comparison schools.
[26] For a detailed discussion, see Appendix Note 1.

equation by Ordinary Least Squares (OLS).

[1] $$Y_{ich} = \beta_1 X_i + \beta_2 A_i + \sum_{k=1} \mu_k I_{ITT\ year=k} + \theta_h + \theta_c + \varepsilon_{ich}$$

In [1], Y_{ich} is the outcome of student *i* in tenth-grade cohort *c*, from high school *h*. X_i is a *matrix* of student demographic characteristics, such as race, gender, and free-lunch status in tenth grade. A_i is a vector of student achievement scores from tenth grade. To control for differences in student attributes across high schools, changes in performance over time, and differences in outcomes across cohorts, I include high school-fixed effects θ_h, and cohort fixed effects θ_c. The variable of interest $I_{ITT\ year=k}$ is an indicator variable denoting the ITT year, so that μ_1 is the effect of the APIP in its first intention-to-treat year, and μ_k is the effect of the APIP in its kth intention-to-treat year. Standard errors are adjusted for clustering at the school level.

A simple before/after comparison likely understates the full effect of the APIP because the first affected cohort is exposed to the two-year program for only one year. As such, the first cohort to receive the "full treatment" is the second cohort (ITT_{year}=2). Because there are likely learning-by-doing effects, and it may take time for the AP program to mature, outcomes may continue to improve beyond the second cohort. As such, to identify the dynamic APIP effect, I use binary variables denoting the first, second, third, and fourth plus ITT years. For example, the first ITT cohort for a school has ITT year=1, and the third ITT cohort for a school has ITT year=3. In this regression, the ITT year denotes how long the APIP had been in place when the student was expected to graduate from high school.[27]

V. Results

<u>*Effects on Short-Run Advanced Placement Outcomes*</u>

While the main objective of this paper is to analyze the long-run effects of the APIP, it is helpful to establish that the program did increase student exposure to the AP program. I present some visual evidence of an APIP effect. Figure 2 shows the results of estimating a flexible version of equation [1], where I estimate effects for both pre-adoption years and post-adoption years. For each outcome, I plot the estimated coefficients of ITT years -4 through 4 (the first

[27] For example, if the APIP were adopted in school *h* in the 2002-03 school year, the tenth-grade cohort for the school year 2000-01 would be coded as ITT year=1, while the tenth-grade cohort for the school year 2002-03 would be coded as ITT year=3.

adoption cohort is year 0 and the first "fully treated" cohort is year 1). There are visible increases in AP course taking and AP exam passing after APIP adoption. These figures also reveal the existence of possible pre-existing upward trends that may lead one to overstate the effects of the APIP. To test for trending formally, I evaluate the null hypothesis that the pre-adoption year effects differ from that of the year immediately preceding adoption, and I fail to reject at the five-percent level *for both outcomes*. Because the statistical tests suggest that a simple before-versus-after comparison might be valid, and because a model with school intercepts and linear time trends is very demanding of the data, the main specifications are based on equation [1]. However, to assuage any lingering concerns about trending, as a robustness check, I present estimates that also include school-specific linear trends in Section V.2.

Table 3 presents the regression results for AP outcomes and college enrollment. In the top panel, I report the coefficient on the simple before/after adoption indicator, and in the lower panel, I report the coefficients on the first, second, third, and fourth plus intention-to-treat years. For AP course taking (column 1), the simple before/after comparison yields a statistically insignificant increase of 0.061. However, the dynamic effect on the lower panel shows a statistically significant increase of 0.165 (about a 21-percent increase) by year three, and the null hypothesis that all the dynamic APIP effects are zero is rejected at the five-percent level. Columns 2 and 3 show the effects on AP exams taken and AP exams passed. The adoption indicator and all the ITT treatment year dummies are statistically significant, and by the fourth year, the APIP is associated with a 0.098 increase (about a 100-percent increase) in the number of AP exams taken, and a 0.043 increase (about a 45-percent increase) in the number of AP exams passed. The increases are primarily in the English and science AP exams (Appendix Table 4). The APIP may affect unmeasured outcomes (such as aspirations or self-confidence) that could in turn impact college outcomes but may not be reflected in these AP outcomes. As such, while these AP effects are important, they may not measure all of the APIP effects and should therefore not be used to scale the effects on college outcomes as if it were a "first stage."

Effects on Medium-Run College Attendance

Before presenting the effect of the APIP on the main long-run outcomes of interest, it is instructive to show that the APIP had the expected effects on medium-run outcomes, such as college enrollment. Figure 3 presents some visual evidence of an APIP effect on college outcomes, where I estimate a flexible version of equation [1], with estimates for both pre-adoption years and post-adoption years (as in Figure 2). On the left panel, there is evidence of a pre-existing upward trend for college attendance, such that it is unclear whether the APIP increased freshman-year college enrollment relative to trend. Indeed, the null hypothesis that the pre-adoption year effects differ from that of the year immediately preceding adoption year can be rejected at the 10-percent level. This suggests that not accounting for pre-existing trends for this particular outcome *may* overstate any effects of the APIP and should be interpreted in that light. As a robustness check, I do present estimates that also include school-specific linear trends.

Columns 4 through 9 of Table 3 present the freshman-year college enrollment effects. Columns 4, 5, 6, and 7 show the effect of the APIP on the likelihood of ever being a freshman at any college, a two-year college, a four-year college, and a private college in Texas, respectively. The before/after comparisons show that the APIP increases college attendance by about 4.2 percentage points on average, and that this is driven by both two-year and four-year college attendance. Because the private colleges are four-year institutions, these results also suggest that a sizable portion of the increase in four-year college attendance occurred at private schools. The dynamic effects demonstrate that the effect of the APIP increases over time, so that by the fourth year of the APIP, college attendance overall increases by 4.8 percentage points (an eight-percent increase); four-year college attendance increases by 2.8 percentage points (a 16-percent increase); and two-year college attendance by a statistically insignificant two percentage points (a five-percent increase). In Section V.2, I show that the APIP exerts positive enrollment effects (although much less precisely estimated) even after accounting for pre-existing trends.

It is important to ensure that the increased college enrollment rates do not reflect shifting college attendance from out-of-state to in-state. I test this directly using national college enrollment data from the NSC for the tenth-grade cohorts of 2005 and 2006. Because I only have NSC data for two cohorts, I can only estimate effects for the first and second ITT cohorts. The results in columns 8 and 9 show that by the second ITT year, the APIP increases in-state college attendance by 3.5 percentage points (significant at the five-percent level), but has no impact on out-of-state college attendance (a point estimate of -0.0001 that is not close to statistical significance in the second ITT year). These results are similar to those obtained using the THECB data (for all years and also for these two cohorts), thus validating the college attendance numbers from the THECB data and showing that the enrollment effects measured in this study reflect real increases in overall college attendance.

Effects on Outcomes at College

Most college attrition occurs in the first year, and thus, persistence through the first year is a key predictor of college success. In the middle panels of Figure 3, I present the dynamic treatment effect on sophomore-year persistence for the four pre-adoption cohorts and five post-adoption cohorts (as with the previous outcomes). While there is minimal evidence of a pre-existing trend for sophomore-year enrollment, there is clear evidence of an increase in this outcome following APIP adoption. I test the null hypothesis that the pre-adoption year effects differ from that of the immediately preceding adoption, and I fail to reject at the 10-percent level, so that trending will not bias estimated effects on this outcome.

The third column of Table 4 presents regression results for sophomore-year college enrollment among those with expected high school graduation by 2006. APIP adoption is associated with a 4.3-percentage point increase in ever being a college sophomore on average, and a 6.6-percentage point increase for the fourth ITT cohort and beyond. This represents a 20-percent increase four years after adoption. Table 4 also shows that affected cohorts are two percentage points more likely to be enrolled in college as a junior, and earn 7.49 more college credits on average (both statistically significant at the five-percent level).

The last college outcome is graduating with a college degree. In the right panel of Figure 3, I present the dynamic treatment effect on earning a college degree. Similar to attending college, there is visible evidence of an upward trend before APIP adoption. I test the null hypothesis that the pre-adoption year effects differ from that of the year immediately preceding adoption, and fail to reject at the five-percent level, but do reject at the 10-percent level; this suggests that not accounting for such trends *may* lead one to overstate the effects of the APIP on college graduation. The regression results in the sixth and seventh columns of Table 4 reveal that despite the improved sophomore and junior year persistence, affected cohorts are no more likely *on average* to earn a college degree. In Sections V.1 and V.2, I show that this overall small effect masks considerable heterogeneity across groups of students.

Effects on Long-Run Labor Market Outcomes

Given that the end goal of all college-preparatory programs is to improve the economic well-being of affected students, the most important outcomes are labor force participation and earnings. These outcomes are particularly interesting because labor market outcomes are measured in 2010, a year in which unemployment was particularly high and in which any differences in human capital should be most pronounced in labor market outcomes. Figure 4 presents the kernel density plot of earnings for the treated and untreated cohorts after removing school-fixed effects and cohort-fixed effects. As one can see, the distribution of residual log earnings is higher among those in the treated cohorts than those in the untreated cohorts. The shift occurs primarily in the middle of the earnings distribution, with no visible effect at the bottom or top of the earnings distribution. While showing that earnings are higher (after accounting for age and school attended) for treated cohorts is important, it is also important to establish that the improved earnings occur after APIP adoption rather than before.

In Figure 5, I present the dynamic treatment effect on the likelihood of having any earnings in 2010 (my proxy for labor market employment) and the log of earnings during the fourth quarter of 2010. While the likelihood of being employed in 2010 is clearly higher after APIP adoption than before, it is clear that some pre-existing trends may account for this. I test the null hypothesis that the pre-adoption year effects differ from those of the first pre-adoption year, and *for this outcome*, I reject at the 10-percent level, suggesting that trending may cause bias in the estimated effect on labor market participation and that these effects should be interpreted with caution. In contrast, there is a clear increase in earnings in 2010 associated with APIP adoption, and there is little visual evidence of trending for this outcome. Consistent with this visual evidence, one cannot the reject the null hypothesis that the pre-treatment year effects differ from those of the year prior to adoption at the 10-percent level, while one rejects the null hypothesis that the post-treatment years differ from the year prior to adoption at the one-percent level. This suggests that positive APIP effects on earning are not the result of pre-existing trends.

The eighth and ninth columns of Table 4 present the regression results for working in 2010 and earnings in 2010. The before-versus-after results indicate that affected students are 2.2 percentage points more likely to be employed than their untreated counterparts, and that this effect increases to 3.8 percentage points for the fourth post-adoption cohort. However, this may be due to trending, so this result should be viewed with some caution. The regression results also suggest effects on earnings conditional on working. Specifically, adoption is associated with 2.7-percent higher earnings on average, and this effect is largest for the second post-adoption cohort, which experiences a 3.7-percent increase in earnings.

V.1 *Effects on Likely AP Course Takers*

The results thus far indicate that the APIP increases AP participation, college attendance, sophomore-year persistence, labor market participation, and earnings. Given that the APIP aims to improve outcomes by improving the quality of AP instruction, increasing student and teacher effort in AP courses, and increasing participation in the AP program, one would expect larger effects for students who are likely to take AP courses. To test this, I estimate the propensity to take AP courses based on observable student covariates and their interactions using the untreated observations, and then I use this model to predict *ex-ante* propensities for all students in the data. Next, I estimate the main treatment effects on the sample of students with estimated propensities above 0.66. If the improved outcomes work through improvement in the AP program (as hypothesized), then the treatment effects should be larger among these students.

I present the simple before/after DID estimates in the top row of Table 5. Consistent with the proposed mechanisms, the treatment effects are much larger for this group of students. Specifically, within this sample, the APIP increases the number of AP exams passed by 0.17 (statistically significant at the 10-percent level), increases college enrollment by 0.08 percentage points, and increases enrolling in college as a sophomore by 9.2 percentage points (both significant at the one-percent level). Unlike with the full sample, these likely AP takers are 1.8 percentage points more likely to graduate from college with any degree. Finally, these students are 2.7 percentage points more likely to work in 2010 and to earn 5.7-percent higher wages. In sum, among students at APIP-adopting schools who are most likely to be affected by the APIP, program adoption is associated with markedly improved educational outcomes and labor market outcomes. This suggests a true causal effect that works through the hypothesized mechanisms. I discuss the changes for those who are unlikely to have been affected in Section V.2.

V.2 *Addressing Threats to Validity*

Because the results are not based on a randomized controlled trial, there is no "silver bullet" identification strategy. As such, while I am careful to compare cohorts within the same school to avoid self-selection within a cohort and selection across schools, and I limit the estimation sample to only the APIP schools that are of similar motivation, a few endogeneity concerns remain. I outline these concerns and present empirical tests to address each concern in this section. While none of these tests is dispositive in isolation, all these tests together present a compelling case that the estimated APIP effects can be interpreted causally.

(a) The timing of APIP adoption may be endogenous. There is the concern that schools that had an *increase* in motivation were more likely to apply to have the APIP implemented. The

APIP takes about two full years to be implemented after a school expresses interest. As such, if the results merely reflected changes in school motivation that coincided with expressing interest in the APIP, one should see an improvement in outcomes two years prior to adoption. While some of the outcomes do exhibit some pre-treatment trends, there is no visual evidence of improvement in outcomes two years before APIP adoption. As a direct test of the "timing of interest" hypothesis, for the main labor market outcomes, I regressed the outcomes on the two-year lag of adoption. This yielded small coefficient estimates and *p*-values larger than 0.05 for all outcomes, and *p*-values larger than 0.1 for the long-run labor market outcomes.

(b) ***The timing of adoption may be associated with the arrival of a new principal.*** Related to the previous point, one may worry that the timing of adoption is related to the arrival of a new principal who makes a variety of new changes *other than the APIP*, which would confound the estimated APIP effects. To test for this, I predict having a new high school principal as a function of whether the program will be adopted in three years, two years, or one year, or was adopted in the same year. These models include school-fixed effects and year-fixed effects only. In each of these four regressions (shown in Appendix Table A1), the *p*-values associated with the null hypothesis of no systematic relationship are larger than 0.2. I also estimate a specification similar to equation (1) and find no effect on subsequent principal turnover (Appendix Table A2). This is consistent with assertions that timing of adoption is idiosyncratic, and suggests that adoption is likely exogenous to *changes* in schools over time.

(c) ***The benefits of the APIP could be driven by general improvements in schools.*** One may wonder whether the benefits of the APIP are driven by other school-wide changes (such as better inputs or change in teaching philosophy) that would operate through some channel other than the AP program. Table 5 shows that the effects are larger for those students with high likelihoods of taking AP courses. If the overall improved effects are truly caused by the AP program, then there should be very small effects for students who are unlikely to take AP courses. If other school-wide policies could have impacted long-run outcomes, then one should see APIP adoption effects on these students. Estimated effects for this sample of students, who are unlikely to take AP courses, are presented in the second panel of Table 5. As one can see, one cannot reject the null hypothesis of no adoption effect for those with estimated likelihood below 0.33 at the 10-percent level. This suggests that AP students receive the benefits of the APIP, and the estimated effects are *not* driven by other confounding changes at schools.

It is important to note that this addresses the issue of pre-existing trends. If the schools are on a trajectory of improvement before APIP adoption, and these trends are responsible for the positive estimated APIP effects, then such trends should be reflected in positive adoption effects for all students. However, across all outcomes, positive effects are only observed among students who are likely to take AP courses. I address the issue of trends further below.

As an additional test, I investigate the APIP effect on school-level outcomes that should not be affected by the APIP as a falsification exercise. Because students who take AP courses are not likely to be on the margin of graduating from high school, there should be no effect of the APIP on high school graduation. Appendix Table 2 shows that APIP adoption is unrelated to graduating from high school, despite the increases in college attendance. Similarly, Appendix Table 2 shows that the APIP has no effect on teacher turnover, total school expenditures, the number of teachers, teacher experience, or average class size.[28] By the fourth year, the only outcome for which there is a statistically significant effect is the number of AP teachers, which is consistent with an expansion of the AP program. For all other outcomes, *including total school spending*, there is no significant effect, suggesting that the effects are not due to general improvements in schools.

(d) ***The estimated APIP effects are due to pre-existing trends.*** Because the visual evidence shows evidence of some trending for some of the outcomes (but not all), there is the concern that the estimated benefits associated with APIP adoption for some outcomes may be confounded with pre-existing upward trends. Even though the test above suggests that the results are probably not driven by pre-existing school-level trends, this is not a direct test. While estimating a model with school fixed effects, cohort fixed effects, and linear time-trends for each school is demanding on the data, it is instructive to see that the main results are generally robust to explicitly accounting for pre-existing time trends. I augment the main estimation model for the main outcomes with cohort fixed effects and both a high school-specific intercept and a linear time trend for each high school (Table 6).[29] While less precise, the estimated effects with linear

[28]While there is little evidence of this, in principle, some of the effects could have been driven by changes in the composition of teachers, as documented in (Jackson, 2009).

[29] That is, I estimate equation [2] below, where all variables are defined as before, $Year_c$ is the cohort's tenth-grade year, and τ_h is a school-specific linear trend for school h. In this model, the APIP effect is the average change in outcomes across treated schools estimated relative to each school's own intercept and own time trend.

[2] $$Y_{ich} = \beta_1 X_i + \beta_2 A_i + \sum_{k=1} \mu_k I_{ITT\ year=k} + \theta_h + \tau_h Year_c + \theta_c + \varepsilon_{ich}$$

trends are similar to the main estimates presented previously. Specifically, *relative to trend*, APIP adoption is associated with statistically significant increases in AP course taking, AP exam taking, AP exam passing, college attendance, sophomore-year college enrollment, employment in 2010, and earnings in 2010. Moreover, the point estimates are similar with the inclusion of trends, so that pre-treatment trending does not explain the estimated effects.

(e) *High-ability, motivated students may self-select into APIP schools after adoption.* Another concern is that these improvements are the result of motivated students self-selecting into secondary schools that adopt the APIP.[30] If positive selection were driving the results, then the APIP should be associated with characteristics that lead to better outcomes. To test for this, I predict the main outcomes as a function of observable student characteristics *before* APIP adoption.[31] I then regress the predicted outcomes on the adoption indicator variables, school effects, and year effects. If there were selection on observable characteristics, APIP adoption would impact these *predicted* outcomes. For all the predicted outcomes, the point estimates are close to zero, and the *p*-values associated with the hypothesis that the adoption year variables are correlated with the predicted outcomes exceed 0.2 (Table 7).[32] This suggests that there is no selection in observed dimensions that are associated with the outcomes.

(f) *There may be selection to APIP schools in unobserved dimensions.* To ensure the results are not driven by selective migration on unobservable characteristics, I estimate equation [1] while including indicator variables for each middle-school-by-high-school combination. Students who self-select into high school because of the APIP will come from middle schools that are not the natural feeder middle schools for the APIP schools (if they were, there would be no need to select). I can avoid comparing the outcomes of students who do self-select to APIP schools from nonfeeder middle schools to those of students who attended the natural feeder middle schools

[30] This is a potential problem because there is the possibility of selective migration. While none of the treated districts allow students from outside the district to enroll, the large urban school districts in Texas (Dallas, Houston, and Austin) practice intradistrict choice, where students have the option to attend their neighborhood schools or another school in the district (including magnet and charter schools), subject to space limitations at the receiving school. Because Houston and Austin schools did not adopt the APIP until 2007, this only poses a problem for Dallas schools. To further ensure that selective migration does not drive the results, I estimate the APIP effect without Dallas schools, and the results are similar. It is also worth noting that under No Child Left Behind, students attending a Title I school designated as "in need of improvement" have the right to attend a higher-performing school in the district. However, in most districts, the APIP schools are the lower-performing schools.

[31] That is, I estimate a regression using the untreated cohort to predict the main outcomes based on all the pre-eleventh-grade covariates and their interactions and then use this model to create predicted outcomes for all students.

[32] I present effects on individual covariates in columns 1 through 6, which indicate that treated cohorts had lower tenth-grade test scores, were less likely to be from low-income households, and were less likely to be Hispanic. While these differences exist, overall treated students had very similar predicted outcomes as untreated students.

and did not self-select by making inferences based on the within-middle-school-by-high-school variation. That is, I only compare the outcomes of students who attended the same middle school and the same high school, so that variation in treatment cannot arise from differences in students' potentially endogenous choice of school. Furthermore, I remove all students who attended middle schools that sent fewer than 300 students to any given APIP high school during the sample period. This removes almost all potential for bias from student selection. The results (Table 6) are similar to the main results, which indicates no selection.[33]

(g) ***Increased college attendance could be due to changes in timing.*** Even though I show that increased college attendance rates are not due to shifting across geographic areas from out-of-state to in-state, readers may wonder whether the enrollment effects merely reflect that the APIP causes students to enroll in school sooner. To assess this, I analyze enrollment within one, two, three, and four years of expected high school graduation in Appendix Table A3. If the effects were due to students enrolling in school sooner rather than later, one would see stronger effects on enrollment close to high school graduation and no effect within longer time horizons. For example, if students attended college within one year of high school graduation rather than two or three years, one would see effects on "ever a freshman within one year of high school graduation" but would see no effect on "ever a freshman within four years of high school graduation." The effects on freshman enrollment and sophomore enrollment grow as one considers longer time horizons, which is exactly what one would see if the effect reflects increased college attendance overall; thus, it is inconsistent with the results being caused by shifting to earlier college entry.

This section shows that the timing of APIP adoption coincides with the timing of improved outcomes; improvements are concentrated among students who should have been affected through the hypothesized mechanisms; there are no impacts on students who are not affected by the hypothesized mechanisms; there are improvements in those outcomes that should be affected by the APIP; there are no improvements in outcomes that should not have been affected by the APIP; the timing of adoption is unrelated to changes in school leadership or

[33] In principle, one could estimate an intention to treat model based on a student's enrollment in middle school. However, estimating such a model in practice becomes complicated because some middle schools feed into multiple high schools, and some high schools draw students from multiple middle schools. Also, this specification would require very large treatment effect among eleventh and twelfth grade students to be detected among eight graders. Furthermore, this would require observing many students five years before high school graduation and would therefore preclude analysis of the first 10 treated schools.

changes in other school inputs; and the effects are not driven by pre-existing improvement over time. As such, even though the variation used is imperfect, given the preponderance of evidence supporting a positive APIP effect, it is reasonable to take the estimates as causal.

V.2 *Effects by Gender and Ethnicity.*

Klopfenstein (2004) documents large differences in AP participation across ethnic groups both across and within schools *(even among students with the same incoming test scores)*, so that there may be differences in treatment status by ethnicity. As such, I estimate APIP effects on sub-samples of students based on gender and ethnicity.[34] I report these effects for black, Hispanic, and white students separately in Table 8. Summary statistics by ethnicity are presented in Appendix Table A4. Some notable differences are observed across ethnic groups. First, while there are small increases in college attendance among black and Hispanic students, white students experience a 7.4-percentage point increase in college enrollment as a result of the APIP. However, all three groups experience sizable increases in sophomore-year college attendance. Specifically, white students experience an 8.9-percentage point increase by the fourth treated cohort, while black and Hispanic students experience 4.7 and 6.2-percentage point increases, respectively. These increases represent about a 10-percent increase for all groups. All groups experience an increase in junior year college attendance and total credits earned. However, Hispanic students experience a 2.6-percentage point increase in BA degree receipt, white students experience no increase, and some evidence suggests a small increase for black students.

While there are no statistically significant increases in labor market participation for black or Hispanic students, there is a 3.7-percentage point increase in the likelihood of being employed in 2010 for white students. However, all groups experience increased earnings. To assuage concerns that these earnings effects are driven by pre-existing trends, I estimate the wage effects in the model with the addition of school-specific linear time trends. In such models, wages for black students increase by 8.2 percent, those for Hispanic students by 5.1 percent, and those for whites by 8.4 percent either three or four years after APIP adoption.

These increased earnings could be due to improved labor market skills after graduating from high school or may reflect increased earnings due to attending college. If these increased earnings were due to increased college attendance, then one should see relatively large wage effects for Hispanic students (for whom there is a clear increase in BA degree receipt) from the

[34] I also estimate effects by gender and find little difference.

early cohorts (who are old enough to have graduated from college and have some labor market earnings). To test for this, I estimate the wage effect using only the cohort of students who were expected to graduate from high school before 2003 while accounting for pre-existing trends. These results in column 12 are consistent with college degree effects driving the wage effects for Hispanic students. Among the older cohorts that have completed their schooling, wage effects (a 14-percent increase by the third post-adoption cohort) are only observed for Hispanic students who also experience an increase in the likelihood of earning a BA degree. The flipside of this result is that improved labor market outcomes for black and white students from the late cohorts cannot be attributed to earning a BA degree, but must be the result of increased skills or the additional schooling (albeit without the degree). These wage increases of roughly 10 percent are similar to those associated with increased math coursework among those who do not attend college documented by Goodman (2009).

V.3 *Evidence of the Mechanisms*:

In this section, I try to shed light on the underlying mechanisms behind the APIP effect.

Are the improvements driven by increased supply of AP courses and sections? It is natural to wonder if the improved outcomes are merely due to an increase in the availability of AP courses and sections. To address this question, I analyze the APIP effect on schools that had above the median number of AP sections before 1996 (schools that had no statistically significant growth in the number of AP sections after adoption) and on those that had below the median number (schools that experienced a statistically significant 150-percent increase in AP sections offered by the fourth adoption year). If the benefits of the APIP were solely due to an increase in the supply of AP sections, one would expect large effects on high-growth schools and small effects on low-growth schools. However, exactly the opposite is true. There are small improvements in college and labor market outcomes for high-growth schools and large improvements for low-growth schools (Table 5), indicating that AP course supply does not drive the results.

Are schools/teachers/students motivated by the incentives? One of the key differences between the APIP and other college-prep programs is the use of cash incentives for students and teachers. As such, it is important to determine whether this is one of the driving mechanisms of the APIP's success. If better resource utilization (caused by the incentives) drives the success of the APIP, and effort is proportional to the size of the rewards, schools that paid higher-powered incentives would obtain better outcomes than those that did not. To test this, I compare the adoption effect on high-power schools (paid between $101 and $500 per exam) and low-powered schools (paid $100 per exam) in Table 5. While the effect on AP exam passing is similar across these school types (Jackson 2010), for all college outcomes, the effects are more significant on high-power schools. Looking to labor market outcomes, the results suggest that both low-powered and high-powered schools see the benefits of the APIP. Students from lower-powered schools see increased labor market participation by 3.5 percentage points with no effect on wages, while those from higher-powered schools see increased wages by 5.4 percent with no effect on labor market participation. Taken as a whole, these findings suggest that the monetary incentives and the increased effort they induce are an important component of the APIP's success.

How important are the non incentive aspects of the program? The improved curricula in earlier grades, teacher training, curricular oversight, vertical teams, and college counseling could all be partially responsible for the success of the APIP. Because the first cohort is only exposed to the APIP for one year, the first cohort to have the full incentive component of the APIP is the second treatment cohort. As such, if the effects were driven by incentives alone, the effects would be the same in the second year as in all subsequent years. For most outcomes, however, this is not the case, suggesting that learning-by-doing or other components of the APIP, which would take a few years to take effect (such as improvements in earlier grades or changes in norms), are central to the effects. One can test for improvements in earlier grades by examining the incoming tenth-grade test scores. The results in Table 7 show that incoming test scores are slightly *lower* after adoption than before, suggesting that improvements in incoming academic preparedness is not the driving force of the APIP's success. This demonstrates that learning-by-doing and changes in school culture are the likely explanations for enhanced APIP effects over time.

Did information or peer norms play a role? Even though I do not have data on student information or on peer norms, to speak to these issues somewhat, I did obtain qualitative interview evidence from guidance counselors. Guidance counselors at three different APIP high schools indicate that school-wide campaigns were implemented to increase their students' participation in AP courses after APIP adoption. At two of the three high schools, an additional guidance counselor was hired to improve these schools' ability to identify those students who should be encouraged to take AP courses. At all three schools, the guidance counselors were given explicit instructions to identify those students who should be taking AP courses and to encourage AP participation. A large part of this campaign involved providing information.

Guidance counselors and AP teachers promoted the AP program to students who were interested in going to college, citing the scholarships one could earn based on AP scores, the tuition one could save by graduating at an accelerated pace, and the potential increase in high school GPA. Guidance counselors also mentioned a shift in student and teacher attitudes toward AP courses, such that AP courses were no longer considered only for the very brightest students.

The tests above suggest that increased supply of courses is not the driving force behind the AP participation response or improved labor market outcomes. That the effects are more significant among high-incentive schools suggests that providing additional supply and removing barriers to taking AP courses alone would not lead to success, but that increased student and teacher effort is an important component of the program. The fact that the effects grow over time suggests that the program aspects emphasized by guidance counselors—such as information and outreach (that schools will have become more proficient at over time)—are also important. The body of evidence indicates that all aspects of the APIP are important and that providing cash incentives to students or teachers alone, providing teacher training alone, or expanding the AP course offerings alone would not have yielded the same results as the full intervention.

VI. Discussion and Conclusions

Using a carefully selected of group of comparison schools within which APIP adoption is likely exogenous, I find that students who were affected by the APIP were more likely to matriculate in college, more likely to persist beyond their freshman year, and those who were most likely to take AP courses were more likely to graduate from a four-year college. Moreover, affected students were more likely to be employed and to earn higher wages. Because both young and old cohorts enjoyed wage increases, the wage increases are consistent with both increased skills attained while in high school and higher wages associated with increased post-secondary educational attainment. In both cases, the college-preparatory program conferred long-run benefits upon affected students. I present a variety of tests, robustness checks, and falsification exercises that suggest that the estimates are not confounded by student selection, school selection, pre-existing trends, changes in leadership, or other coincident school policies.

The APIP led to larger improvements in educational attainment and earnings for Hispanic students (the group with the lowest baseline college attendance rates) than for white and black students. These findings suggest that, in addition to reducing ethnic gaps within schools, because the program was targeted to inner-city school with low shares of high income and white students, the program also helped to reduce educational and earnings gaps overall. The earnings increases associated with the APIP for Hispanic and black students are large enough to reduce the black-white earnings gap by one third and to eliminate the Hispanic-white earnings gap entirely.

Given that I find no evidence of worse outcomes associated with the APIP, these improvements are likely the result of increased exposure to rigorous. Consistent with this interpretation, APIP adoption is associated with increased AP course taking and AP examination taking, and the effects only exist for students who were *ex ante* likely to take AP courses. The evidence on mechanisms indicates that both the incentive aspects and the non incentive aspects are important. The lack of any documented ill effects of the APIP suggests that many of the hypothesized detrimental effects of using student incentives or teacher performance pay need not pose a large practical problem in a well-designed incentive-based scheme that combines incentives with additional resources to help translate increased effort into results.

The total cost of the program is roughly $225 per high school junior and senior. Because most students are exposed to the program for two years, the average cost per affected student is $450. Using the smallest estimated wage increases as an estimate of the benefits, among students exposed to the APIP for two years, the increase in earnings in 2010 was 3.7 percent. For individuals with baseline annual earnings of $25,000 per year, this would be an annual increase of $925. As such, the benefit-cost ratio for this low level of earnings would be about 2:1 if this was a one-time wage increase. Under the more realistic assumption that the 3.7-percent increase is persistent and affects lifetime earnings, an individual with a starting wage of $25,000 would have an approximate present discounted value of lifetime earnings of $450,000.[35] This would imply a lifetime benefit of the APIP of $16,650 and a benefit-to-cost ratio of 37:1. With higher baseline earnings, this ratio would be even higher.

Because the large increases in AP participation imply that low AP participation may have reflected some sub-optimality such as poor information, sub-optimal peer norms, or barriers to taking AP exams, it is unsurprising that the economic returns to the program are large. The improvements imply that it may be possible to enhance outcomes by improving both students' and teachers' decision-making and increasing access to well-taught rigorous courses.

While recent evidence has demonstrated that moving students out of low-performing

[35] Assuming a working life of 35 years, wage growth of two percent per year and a discount rate of seven percent.

schools and into high-performing schools *can* improve student outcomes, very little evidence has shown that one can improve students' long-run outcomes by adopting a program at their existing schools. Because there has been little credible evidence on the efficacy of college-prep programs despite large public and private expenditure on such programs, the results of this study are encouraging about the potential efficacy of college-preparatory programs at improving the educational outcomes of disadvantaged students who are consigned to inner-city schools.

Bibliography

Abdulkadiroğlu, A., J.D. Angrist, S.M. Dynarski, T.J. Kane, and P.A. Pathal. 2011. "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." *Quarterly Journal of Economics* 126, no. 2:699-748.

Acemoglu, Daron and David H. Autor. 2010. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In Orley Ashenfelter and David Card,eds., *Handbook of Labor Economics,* Elsevier, Vol. 4B, 1043-1171.

Adelman, C. 1999 *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment.* Washington, DC: U.S. Department of Education.

Angrist, J., and V. Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial." *Amercian Economic Review* 99:1384-1414.

Angrist, J., D. Lang, and P. Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 1, no. 1:136-163.

Booker, K., T.R. Sass, B. Gill, and R. Zimmer. 2011. "The Effect of Charter High Schools on Educational Attainment." *Journal of Labor Economics* 29, no. 2:377-415.

Bound, John, Michael F. Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." American Economic Journal: Applied Economics, 2(3): 129–57.

Cahlan, M. 2008. "A PART Tragedy: The Case of Upward Bound Correcting for Study Error in the Random Assignment 1992-2004 National Evaluation of Upward Bound." *United States Department of Education Report*.

Cameron, J., and D.W. Pierce. 1994. "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis." *Review of Educational Research* 64:363-423.

Cunha, F., and J. Heckman. 2007. “The Technology of Skill Formation.” *American Economic Review* 97, no. 2:31-47.

Deci, E. L., and R.M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Deming, D., J. Hastings, T. Kane, and D. Staiger. 2011. “School Choice, School Quality and Academic Achievement.” Unpublished.

Dobbie, W., and R.G. Fryer. 2011. “Living in the Zone: An Analysis of a Bold Social Experiment in Harlem.” *American Economic Journal: Applied Economics* 3, no. 3:158-187.

Ellwood, David and Thomas J. Kane, “Who is Getting a College Education: Family Background and the Growing Gaps in Enrollment” in Sheldon Danziger and Jane Waldfogel (eds.) *Securing the Future* (New York: Russell Sage Foundation, 2000).

Figlio, D. N., and L.W. Kenny. 2007. “Individual Teacher Incentives and Student Performance.” *Journal of Public Economics* 91, nos. 5-6:901-914.

Flores-Lagunes, Alfonso., and Audrey Light, 2010."Interpreting Degree Effects in the Returns to Education," Journal of Human Resources, University of Wisconsin Press, vol. 45(2).

Goodman, J. 2009. “The Labor of Division: Returns to Compulsory Math Coursework.” Working Paper Harvard University.

Hudgins, K. 2003, May. “Advanced Placement Program Proves It Pays to Study Hard: A Kick Start for College.” *Fiscal Notes*.

Jackson, C. K. 2010. “A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program. *Journal of Human Resources* 45, no. 3.

Jackson, C. K. 2010. “Do Students Benefit From Attending Better Schools?: Evidence from Rule-based Student Assignments in Trinidad and Tobago.” *Economic Journal*.

Jackson, C. K. 2009. “Student Demographics, Teacher Sorting, and Teacher Qualty: Evidence from the End of School Desegregation.” *Journal of Labor Economics* 27, no. 2:213-256.

Jackson, C. K. 2010. “The Effects of an Incentive-Based High-School Intervention on College Outcomes.” NBER Working Paper 15722.

Jackson, C. K., and E. Bruegmann. 2009. “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers.” *American Economic Journal: Applied Economics* 1, no. 4:85-108.

Kemple, J. J., and C.J. Willner. 2008. *Career Academies Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood.* MDRC.

Klopfenstein, K. 2004. “The Advanced Placement Expansion of the 1990s: How Did Traditionally Underserved Students Fare?” *Education Policy Analysis Archives* 12, no. 68.

Kohn, A. 1999. *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes.* Bridgewater, NJ: Replica Books.

Lavy, V. 2009. “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics.” *American Economic Review*.

Lerner, J. B., and B. Brand. 2008. “Review of State Policies Supporting Advanced Placement, International Baccalaureate, and Dual Credit Programs.” *Baccalaureate and Dual Credit Programs*.

Mathews, J. 2004. “Paying Teachers and Students for Good Scores.” *The Washington Post*, August 10.

Mervis, J. 2011. “Navy Paying Students to Succeed on AP Tests.” *Education Next*, October 27.

Murnane, R. J. 2008. “Educating Urban Children.” NBER Working Paper 13791.

Neal, D., and W. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *The Journal of Political Economy* 104, no. 5:869-895.

Rouse, Cecilia and Lisa Barrow. 2006. “U.S. Elementary and Secondary Schools: Equalizing Opportunity or Replicating the Status Quo?” *The Future of Children* 16(2): 99-123.

Seftor, N. S., A. Mamun, and A. Schirm. 2009. “The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation.” *Mathematica* Report.

Turner, Sarah. 2004. “Going to College and Finishing College: Explaining Different Educational Outcomes.” In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, edited by Caroline M. Hoxby, 13-61. Chicago: University of Chicago Press.

Figures and Tables

Figure 1: *Number of Schools Adopting APIP: By Year*

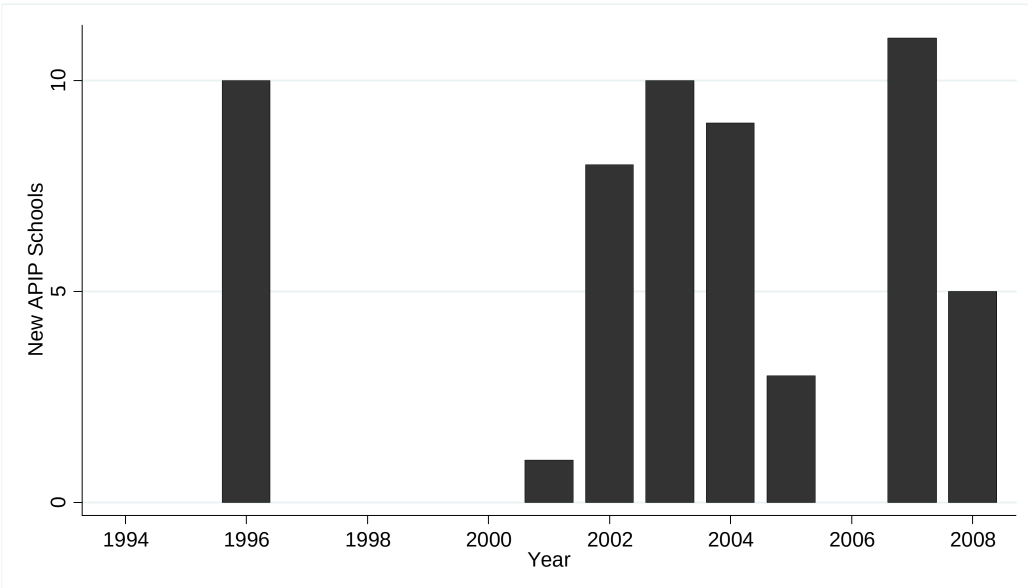
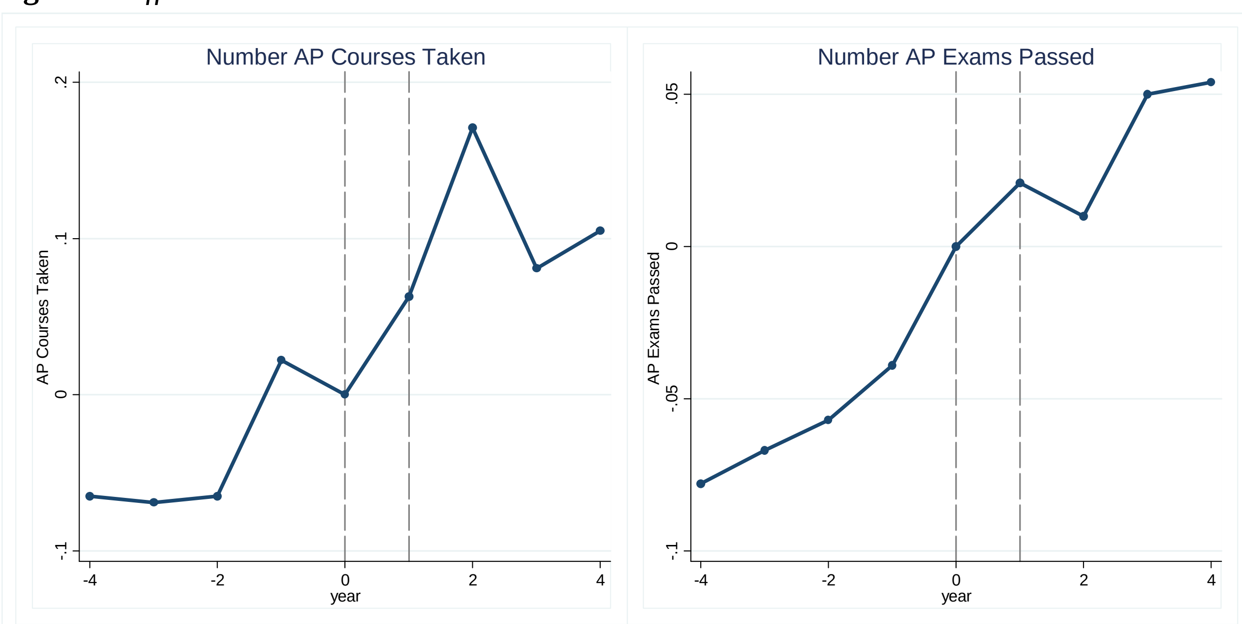
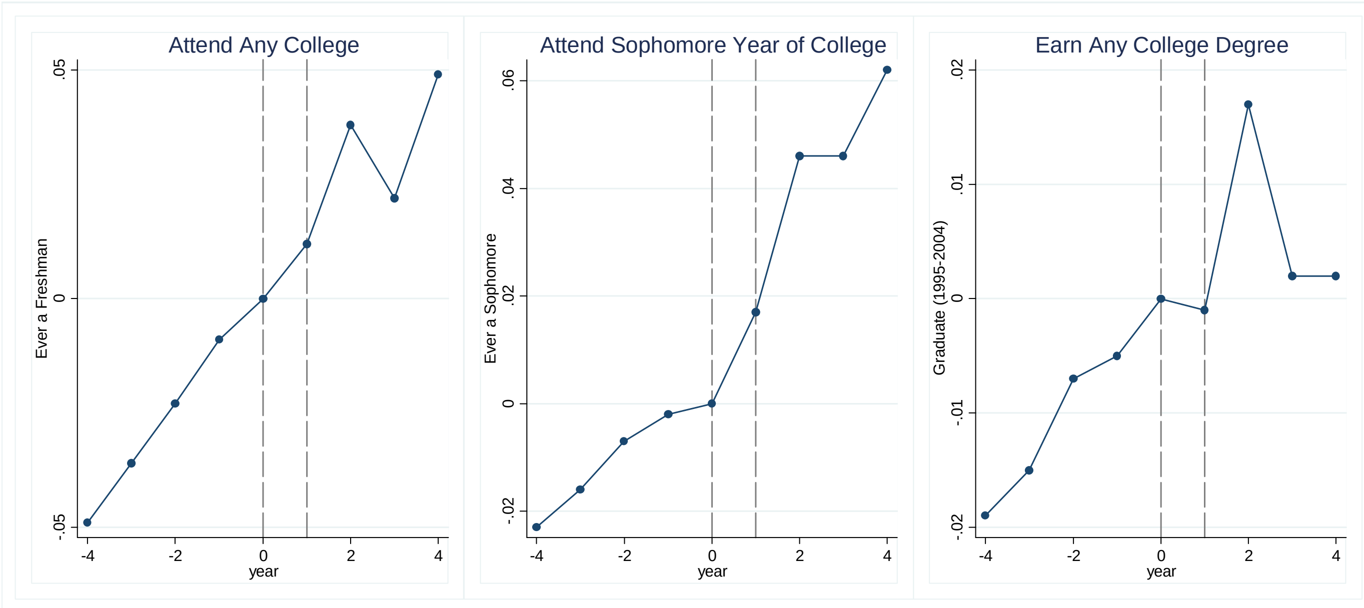


Figure 2: *Effect on AP Outcomes*



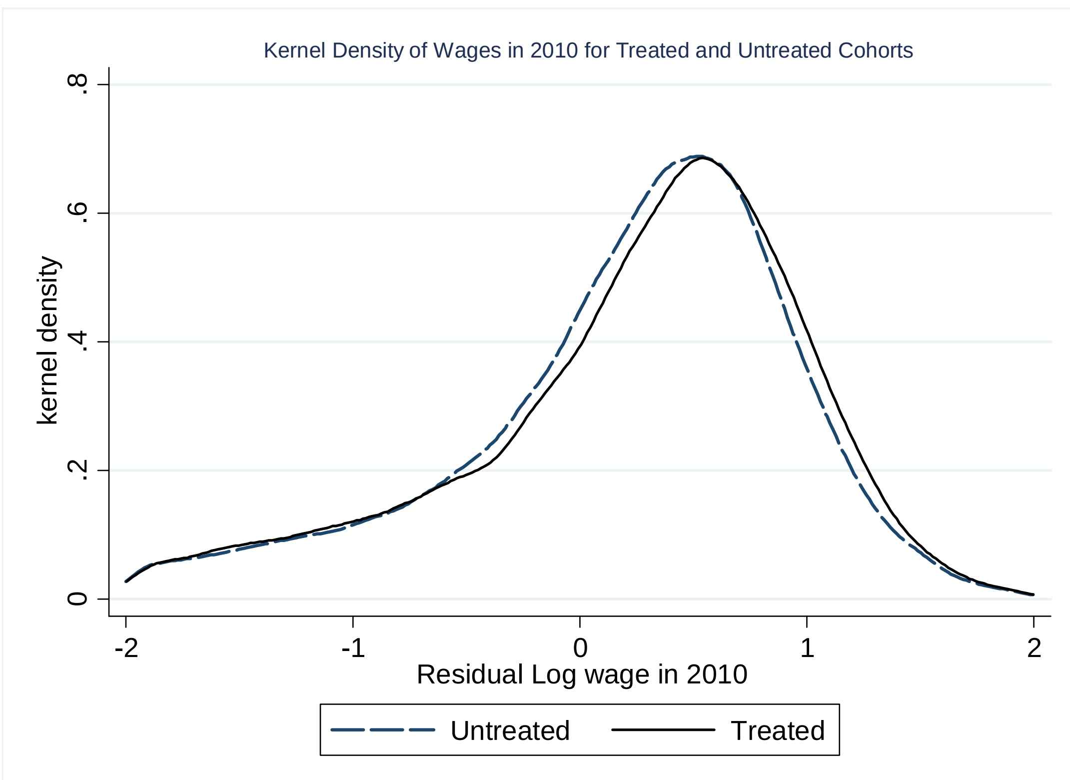
This figure shows the results of estimating a flexible version of equation [1], where I estimate effects for both pre-adoption years and post-adoption years. For each outcome, I plot the estimated coefficients of ITT years -4 through 4 (the first adoption cohort is year 0 and the first "fully treated" cohort is year 1).

Figure 3: *Effect on College Outcomes*



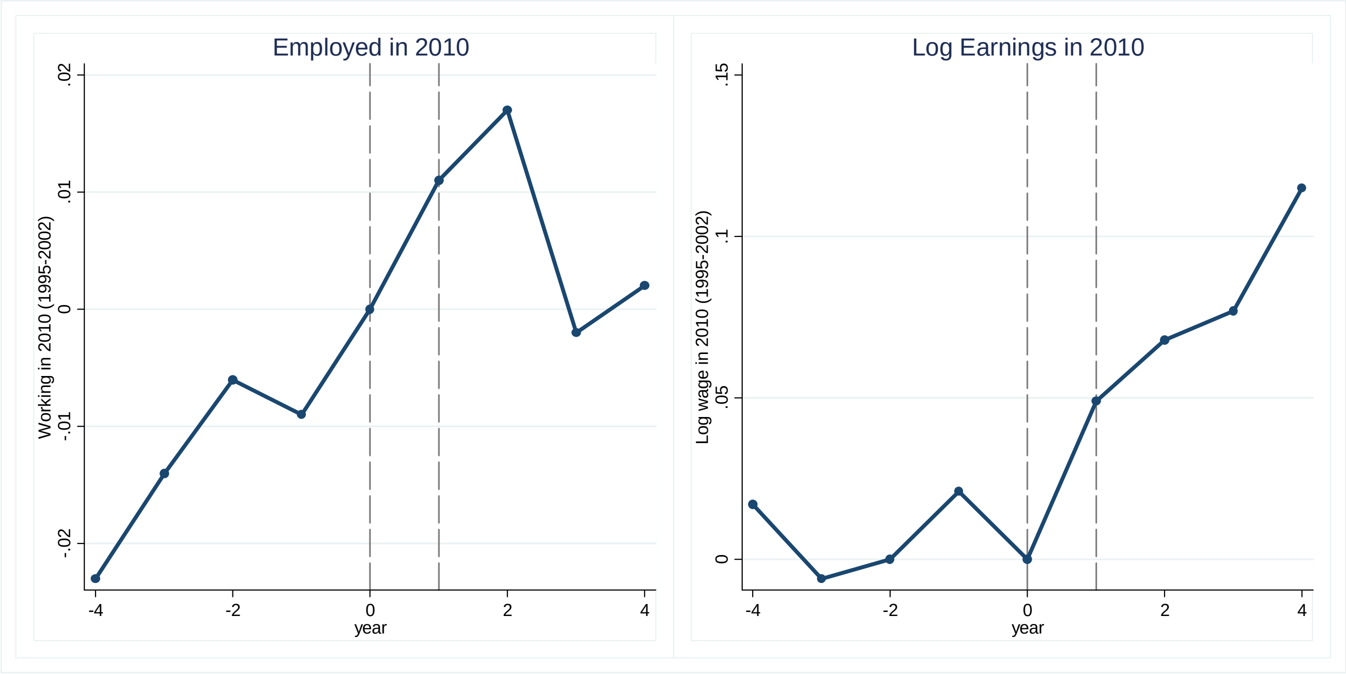
This figure shows the results of estimating a flexible version of equation [1], where I estimate effects for both pre-adoption years and post-adoption years. For each outcome, I plot the estimated coefficients of ITT years -4 through 4 (the first adoption cohort is year 0 and the first "fully treated" cohort is year 1).

Figure 4: *Wages for Treated and Untreated Cohorts*



This figure presents the kernel density plot of earnings for the treated and untreated cohorts after removing school-fixed effects and cohort-fixed effects.

Figure 5: *Effect on Labor Market Outcomes*



This figure shows the results of estimating a flexible version of equation [1], where I estimate effects for both pre-adoption years and post-adoption years. For each outcome, I plot the estimated coefficients of ITT years -4 through 4 (the first adoption cohort is year 0 and the first "fully treated" cohort is year 1).

Table 1

SUMMARY STATISTICS FOR DEMOGRAPHICS FOR APIP SCHOOLS AND OTHER COMPARISON GROUPS

	APIP Schools		Non-APIP Schools	
	1993-1999	2000-2005	1993-1999	2000-2005
Enrollment	1777.68	1836.36	716.85	751.56
	(642.34)	(648.86)	(781.97)	(833.36)
% White	30.82	25.16	59.38	53.36
	(25.43)	(23.28)	(29.46)	(30.42)
% Black	30.17	26.24	10.32	11.30
	(26.82)	(23.5)	(15.64)	(17.08)
% Hispanic	35.76	45.36	28.92	33.67
	(23.49)	(23.84)	(28.9)	(29.5)
% Asian	2.93	2.39	1.09	1.12
	(3.43)	(3.65)	(2.76)	(2.98)
% Free lunch	34.33	41.60	30.42	35.51
	(22.3)	(25.0)	(23.97)	(26.25)
% Limited English	9.66	10.68	3.57	3.83
	(12.89)	(11.86)	(7.71)	(6.8)
City	0.874	0.739	0.182	0.197
	(0.28)	(0.44)	(0.39)	(0.4)
Rural	0.000	0.017	0.489	0.373
	(0.0)	(0.13)	(0.5)	(0.48)
Number of Schools	58		1413	

Standard deviations in parentheses.

Table 2

Student-Level Summary Statistics of APIP Schools Before and After APIP Adoption

Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
	Not Adopted APIP			Adopted APIP		
Grade 10 Year	156858	1998.678	(3.247)	137704	2003.117	(3.335)
LEP	156858	0.112	(0.315)	137704	0.138	(0.345)
Low Income	156858	0.384	(0.486)	137704	0.459	(0.498)
Black	156858	0.206	(0.404)	137704	0.270	(0.444)
Hispanic	156858	0.444	(0.497)	137704	0.426	(0.494)
Asian	156858	0.034	(0.182)	137704	0.037	(0.188)
Native American	156858	0.003	(0.059)	137704	0.004	(0.062)
Female	156858	0.502	(0.5)	137704	0.510	(0.5)
Tenth-Grade Reading z-Score	156858	-0.092	(1.018)	137704	-0.063	(0.987)
Tenth-Grade Math z-score	156858	-0.091	(1.004)	137704	-0.078	(0.962)
Take AP Course	156858	0.229	(0.42)	137704	0.304	(0.46)
AP Courses Taken	156858	0.652	(1.539)	137704	0.974	(1.947)
Take AP Exam	155753	0.055	(0.228)	138535	0.068	(0.252)
AP Exams Taken	155753	0.097	(0.506)	138535	0.127	(0.598)
AP Exams Passed	155753	0.047	(0.342)	138535	0.054	(0.366)
Freshman at Any School	156858	0.592	(0.691)	137704	0.570	(0.684)
Ever Freshman at Private	44157	0.020	(0.142)	111838	0.041	(0.197)
Ever Freshman at Four-year	156858	0.174	(0.386)	137704	0.177	(0.391)
Ever Freshman at Two-year	156858	0.418	(0.493)	137704	0.392	(0.488)
Attend College outside TX	2018	0.045	(0.208)	42293	0.037	(0.189)
Attend College in TX	2018	0.398	(0.49)	42293	0.421	(0.494)
Freshman Year GPA	58685	2.382	(1.176)	44425	2.427	(1.192)
Sophomore at Any School	156858	0.314	(0.561)	115783	0.310	(0.562)
Junior at Any School	154840	0.164	(0.375)	95411	0.155	(0.366)
Graduate with a BA	147779	0.146	(0.354)	80931	0.115	(0.319)
Graduate with an AA	156858	0.038	(0.191)	115783	0.021	(0.142)
Working 2010	156858	0.4925	(0.500)	160776	0.46774	(0.498)
Wage in 2010 (4th quarter)	77263	8029.6	(7990)	75202	4895.2	(4616)

Table 3

Regression Estimates: Effect of Years of APIP Adoption on AP Course Taking and College Enrollment

	1	2	3	4	5	6	7	8	9
	TEA Data			Texas Higher Education Data				NSC Data	
	AP Courses Taken	AP Exams Taken	AP Exams Passed	Ever Freshman	Ever Freshman at Two-yr	Ever Freshman at Four-yr	Ever Freshman at Private School	Enrolled Out-of - State	Enrolled In-State
Adopted	0.061	**0.0708**	**0.0246**	**0.042**	**0.019**	**0.023**	**0.015**	-0.001	0.021
(ITT year>0) years=1	[0.063]	**[0.011]****	**[0.009]****	**[0.014]****	**[0.010]+**	**[0.008]****	**[0.005]****	[0.015]	[0.015]
ITT year=1	-0.008	**0.066**	**0.031**	**0.029**	0.013	**0.016**	**0.011**	-0.0002	0.016
	[0.053]	**[0.012]****	**[0.008]****	**[0.013]***	[0.010]	**[0.007]***	**[0.004]***	[0.0146]	[0.014]
ITT year=2	0.057	**0.086**	**0.042**	**0.043**	**0.02**	**0.023**	**0.014**	-0.001	**0.035**
	[0.063]	**[0.015]****	**[0.008]****	**[0.015]****	**[0.011]+**	**[0.008]****	**[0.005]***	[0.015]	**[0.017]***
ITT year=3	**0.165**	**0.066**	**0.028**	**0.066**	**0.033**	**0.034**	**0.021**	-	-
	[0.080]*	**[0.017]****	**[0.009]****	**[0.018]****	**[0.013]***	**[0.011]****	**[0.007]****	-	-
ITT year=4+	0.074	**0.098**	**0.043**	**0.048**	0.02	**0.028**	**0.018**	-	-
	[0.105]	**[0.021]****	**[0.011]****	**[0.020]***	[0.015]	**[0.014]+**	**[0.007]****	-	-
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FX	YES	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES	YES
F-test	0.011	>0.000	>0.000	0.007	0.087	0.021	0.001	0.81	0.17
Observations	290,343	290,343	290,343	290,343	290,343	290,343	290,343	44,311	44,311

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.

** p<0.01, * p<0.05, + p<0.1

All regressions control for tenth-grade test scores, ethnicity, gender, LEP status, and free or reduced lunch status.

Table 4

APIP Effect on Medium- and Long-run Outcomes

	Ever a freshman	Freshman-year GPA	Ever a sophomore	Ever a Junior	Total Credits	Any degree	Degree within 4 years	Working in 2010	Log Wage in 2010
	Expected High School Graduation Before 2005							All cohorts	
Adopted	**0.042**	**0.04**	**0.043**	**0.02**	**7.497**	0.004	0.002	**0.022**	**0.027**
	[0.014]*	**[0.017]***	**[0.010]****	**[0.007]****	**[1.923]****	[0.011]	[0.003]	**[0.009]***	**[0.016]+**
ITT year= 1	**0.033**	0.009	**0.027**	**0.016**	**4.395**	0.005	0.002	**0.017**	0.018
	[0.017]+	[0.020]	**[0.010]***	**[0.006]****	**[1.843]***	[0.006]	[0.003]	**[0.009]+**	[0.017]
ITT year= 2	**0.053**	**0.046**	**0.033**	**0.019**	**6.626**	0.003	0	**0.027**	**0.037**
	[0.017]**	**[0.026]+**	**[0.010]****	**[0.007]****	**[1.708]****	[0.009]	[0.004]	**[0.010]****	**[0.017]***
ITT year= 3	**0.065**	**0.048**	**0.061**	**0.022**	**9.368**	0.005	0	**0.038**	**0.032**
	[0.019]**	**[0.029]+**	**[0.013]****	**[0.009]***	**[2.205]****	[0.012]	[0.004]	**[0.012]****	**[0.020]+**
ITT year= 4+	**0.04**	**0.066**	**0.065**	0.013	**4.185**	0.018	**0.01**	**0.038**	0.009
	[0.023]+	**[0.029]***	**[0.016]****	[0.011]	**[2.413]+**	[0.016]	**[0.006]+**	**[0.016]***	[0.036]
School FX	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FX	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	224971	93286	224971	224971	224971	224971	224971	290,343	290,343

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.
All regressions control for tenth-grade test scores, gender, ethnicity, LEP status, and free or reduced lunch status.

Table 5

Regression Estimates of the Effect of APIP Adoption Duration on Outcomes: By Sub-sample

	1	2	3	4	5	6	7
	AP Courses Taken	AP Exams Passed	Ever a Freshman	Ever a Sophomore	Graduate from College	Working in 2010	Log wage in 2010
				High Likelihood of Taking AP Courses			
Treated	0.203	**0.179**	**0.08**	**0.092**	**0.018**	**0.027**	**0.057**
	[0.178]	**[0.108]+**	**[0.027]****	**[0.024]****	**[0.011]+**	**[0.013]***	**[0.027]***
				Low Likelihood of Taking AP Courses			
Treated	-0.052	0.003	0.009	0.008	-0.004	0.017	-0.012
	[0.035]	[0.004]	[0.012]	[0.006]	[0.003]	[0.016]	[0.040]
				High AP Course Capacity before Adoption			
Treated	**0.135**	**0.047**	**0.047**	**0.06**	0.005	**0.025**	**0.042**
	[0.064]*	**[0.011]****	**[0.021]***	**[0.012]****	[0.009]	**[0.013]+**	**[0.023]+**
				Low AP Course Capacity before Adoption			
Treated	-0.01	**0.0362**	0.027	-0.009	-0.011	0.016	-0.006
	[0.116]	**[0.008]****	[0.017]	[0.013]	[0.008]	[0.015]	[0.029]
				High-Power Incentives			
Treated	**0.135**	**0.022**	**0.036**	**0.055**	0.011	0.004	**0.054**
	[0.058]*	**[0.009]***	**[0.021]+**	**[0.013]****	[0.011]	[0.014]	**[0.031]+**
				Low-Power Incentives			
Treated	-0.025	**0.027**	**0.044**	0.006	-0.012	**0.035**	-0.006
	[0.087]	**[0.010]***	**[0.021]***	[0.013]	[0.008]	**[0012]****	[0.025]

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

All regressions control for tenth-grade test scores, ethnicity, gender, LEP status, and free or reduced lunch status.

Table 6

Regression Estimates: Effect of APIP Adoption Years on College Outcomes for Different Sub-samples (Robustness Check)

	1	2	3	4	5	6	7	8
	AP Courses Taken	AP Exams Taken	AP Exams Passed	Ever a freshman	Ever sophomore	Graduate with any degree	Working in 2010	Log of wages in 2010
	Middle school- and high school-fixed effects and only feeder schools							
ITT year= 1	**0.096**	**0.075**	**0.039**	0.013	-0.009	0	0.007	0.022
	[0.052]+	**[0.018]****	**[0.012]****	[0.015]	[0.014]	[0.007]	[0.008]	[0.022]
ITT year= 2	**0.166**	**0.085**	**0.046**	**0.037**	0.01	-0.004	**0.021**	**0.037**
	[0.065]*	**[0.021]****	**[0.011]****	**[0.017]***	[0.016]	[0.009]	**[0.010]***	**[0.018]+**
ITT year= 3	**0.3**	**0.086**	**0.041**	**0.054**	**0.047**	-0.001	**0.033**	0.029
	[0.080]**	**[0.023]****	**[0.012]****	**[0.017]****	**[0.014]****	[0.010]	**[0.012]****	[0.028]
ITT year= 4+	**0.172**	**0.122**	**0.06**	**0.052**	**0.047**	-0.032	**0.046**	0.017
	[0.083]*	**[0.031]****	**[0.017]****	**[0.022]***	**[0.018]***	[0.023]	**[0.017]****	[0.036]
	School intercepts and trends included							
ITT year= 1	0.003	**0.065**	**0.034**	0.008	-0.001	0.007	0.006	0.022
	[0.050]	**[0.016]****	**[0.010]****	[0.011]	[0.010]	[0.005]	[0.007]	[0.016]
ITT year= 2	0.074	**0.084**	**0.045**	**0.025**	0.004	0.007	**0.016**	**0.039**
	[0.062]	**[0.016]****	**[0.009]****	**[0.015]+**	[0.012]	[0.005]	**[0.009]+**	**[0.019]***
ITT year= 3	**0.201**	**0.08**	**0.04**	**0.03**	**0.031**	0.008	**0.027**	**0.041**
	[0.082]*	**[0.022]****	**[0.011]****	**[0.017]+**	**[0.014]***	[0.006]	**[0.011]***	**[0.024]+**
ITT year= 4+	0.113	**0.129**	**0.065**	0.017	**0.028**	0.008	0.016	0.01
	[0.105]	**[0.031]****	**[0.017]****	[0.021]	**[0.016]+**	[0.011]	[0.011]	[0.025]

Heteroskedasticity robust standard errors in parentheses are adjusted for clustering at the school level.
+significant at 10 percent; * significant at 5 percent; ** significant at 1 percent. All models include cohort-fixed effects and high school-fixed effects. All models except the top panel include the full set of controls, as in Table 3.

Table 7

Regression Estimates: Effect of APIP Program on Selected Student Characteristics by Adoption Cohort

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Math Score 10th Grade	Reading Score 10th Grade	LEP	Low Income	Black	Hispanic	Predicted: AP Courses	Predicted: AP Exams	Predicted: Attend College	Predicted: Sophomore	Predicted: College Graduate	Predicted: Working in 2010	Predicted: log wages in 2010
Adopted	-0.039	-0.013	0.009	-0.039	0.005	-0.025	0.01	-0.004	-0.001	-0.002	-0.002	-0.004	0.001
ITT year>0	[0.025]	[0.021]	[0.011]	[0.018]*	[0.011]	[0.012]*	[0.032]	[0.040]	[0.006]	[0.007]	[0.003]	[0.003]	[0.007]
ITT year=1	-0.02	0.013	0.009	-0.021	0.003	-0.014	0.002	-0.007	-0.003	0	-0.002	-0.004	0.001
	[0.026]	[0.021]	[0.007]	[0.015]	[0.008]	[0.010]	[0.026]	[0.024]	[0.005]	[0.006]	[0.003]	[0.003]	[0.007]
ITT year=2	**-0.051**	-0.035	0.005	**-0.044**	0.005	**-0.024**	0.015	0.004	-0.005	0.005	-0.002	-0.004	0.001
	[0.032]+	[0.024]	[0.012]	**[0.021]***	[0.011]	**[0.013]+**	[0.038]	[0.045]	[0.006]	[0.008]	[0.003]	[0.003]	[0.007]
ITT year=3	-0.049	-0.016	0.011	**-0.049**	0.001	**-0.04**	0.014	-0.024	-0.005	0.005	-0.002	-0.004	0.001
	[0.034]	[0.030]	[0.016]	**[0.024]***	[0.015]	**[0.017]***	[0.041]	[0.039]	[0.006]	[0.008]	[0.003]	[0.003]	[0.007]
ITT year=4+	-0.05	-0.047	0.017	**-0.068**	0.012	**-0.045**	0.018	0.017	-0.007	-0.002	-0.002	-0.004	0.001
	[0.033]	[0.29]	[0.022]	**[0.030]***	[0.016]	**[0.020]***	[0.048]	[0.052]	[0.007]	[0.010]	[0.003]	[0.003]	[0.007]
School FX	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FX	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	290343	290343	290343	290343	290343	290343	290343	290343	290343	290343	290343	290343	290343

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.

+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

All regressions control for tenth-grade test scores, ethnicity, gender, LEP status, and free or reduced lunch status.

Table 8

Regression Estimates: Effect of Years of APIP Adoption on Longer-run College Outcomes by Ethnicity

	1	2	3	4	5	6	7	8	9	12
	DID	DID	DID	DID	DID	DID	DID	DID	DID with Trends	DID with Trends
	Ever a freshman	Ever a sophomore	Ever a Junior	Total Credits	Any degree	BA	Working in 2010 (all cohorts)	Log Wage in 2010 (all cohorts)	Log Wage in 2010 (all cohorts)	Log Wage in 2010 (early cohorts)
	Black (54059 observations)									
ITT year= 1	0.001	0.011	**0.016**	3.131	0.016	**0.018**	0.013	-0.007	0.003	-0.082
	[0.026]	[0.014]	**[0.009]+**	[2.819]	[0.010]	**[0.009]***	[0.017]	[0.036]	[0.027]	[0.056]
ITT year= 2	0.013	0.011	0.011	3.461	0.01	0.011	0.018	0.069	0.024	-0.003
	[0.022]	[0.013]	[0.008]	[2.247]	[0.011]	[0.011]	[0.014]	[0.052]	[0.034]	[0.055]
ITT year= 3	**0.038**	**0.034**	**0.017**	**7.68**	0.017	**0.018**	0.018	**0.093**	**0.054**	0.009
	[0.021]+	**[0.014]***	**[0.010]+**	**[2.600]****	[0.012]	**[0.011]+**	[0.019]	**[0.051]+**	**[0.026]***	[0.056]
ITT year= 4+	-0.012	**0.047**	0.012	5.997	0.003	0.01	0.01	**0.145**	**0.082**	0.01
	[0.025]	**[0.020]***	[0.012]	[3.674]	[0.012]	[0.011]	[0.022]	**[0.041]****	**[0.046]+**	[0.044]
	Hispanic (93171)									
ITT year= 1	0.015	0.021	**0.013**	1.68	**0.011**	**0.012**	0.01	**0.043**	0.033	0.047
	[0.016]	[0.013]	**[0.006]***	[2.025]	**[0.006]+**	**[0.005]***	[0.012]	**[0.022]+**	[0.026]	[0.043]
ITT year= 2	0.03	**0.035**	**0.022**	**4.561**	**0.018**	**0.022**	0.015	0.033	0.023	0.072
	[0.018]	**[0.014]***	**[0.008]****	**[2.372]+**	**[0.010]+**	**[0.008]***	[0.014]	[0.024]	[0.045]	[0.046]
ITT year= 3	0.027	**0.052**	**0.021**	**6.828**	**0.024**	**0.026**	0.014	0.038	**0.042**	**0.14**
	[0.021]	**[0.014]****	**[0.007]****	**[1.949]****	**[0.010]***	**[0.010]***	[0.016]	[0.030]	**[0.024]+**	**[0.050]****
ITT year= 4+	0.016	**0.062**	**0.021**	4.793	**0.023**	**0.025**	0.021	**0.081**	**0.051**	**0.117**
	[0.028]	**[0.019]****	**[0.009]***	[2.878]	**[0.012]+**	**[0.011]***	[0.024]	**[0.033]***	**[0.027]+**	**[0.056]***
	White (68722)									
ITT year= 1	**0.041**	**0.038**	**0.024**	4.909	0.004	0.01	0.011	**0.044**	**0.041**	-0.019
	[0.024]+	**[0.015]***	**[0.011]***	[3.004]	[0.009]	[0.009]	[0.012]	**[0.026]+**	**[0.024]+**	[0.032]
ITT year= 2	**0.064**	**0.042**	**0.024**	**6.505**	0.004	0.013	0.019	**0.109**	**0.077**	-0.01
	[0.025]*	**[0.017]***	**[0.012]+**	**[2.679]***	[0.011]	[0.012]	[0.012]	**[0.035]****	**[0.037]***	[0.059]
ITT year= 3	**0.078**	**0.075**	**0.027**	**9.038**	-0.005	0.005	**0.041**	**0.074**	**0.084**	-0.062
	[0.029]**	**[0.021]****	**[0.016]+**	**[2.763]****	[0.012]	[0.013]	**[0.015]****	**[0.040]+**	**[0.040]***	[0.078]
ITT year= 4+	**0.074**	**0.089**	0.024	**7.107**	-0.017	-0.011	**0.037**	0.032	0.047	-0.038
	[0.034]*	**[0.023]****	[0.017]	**[2.815]***	[0.018]	[0.018]	**[0.016]***	[0.045]	[0.052]	[0.080]

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.
All regressions control for tenth-grade test scores, gender, ethnicity, LEP status, and free or reduced lunch status.

Appendix

Appendix Table A1: *Principal Turnover and APIP adoption*

	1	2	3	4	5	6
	Principal turnover(t-3)	Principal turnover(t-2)	Principal turnover(t-1)	Principal turnover(t)	Principal turnover(t+1)	Principal turnover(t+2)
Adopted	0.004	0.058	0.003	0.026	0.004	-0.092
ITT year>0	[0.059]	[0.066]	[0.046]	[0.055]	[0.068]	[0.070]
Observations	411	466	473	583	530	580

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
* significant at 5 percent; ** significant at 1 percent. All regressions include school- and year-fixed effects.

Appendix Table A2

	1	2	3	4	5	6	7	8	9
	High School Diploma	Teacher Turnover (Proportion)	Principal turnover	Log Expenditure on Instruction	Log Total Expenditure	Number of AP Teachers	Number of Non AP Teachers	Mean Teacher Experience	Mean Class Size
ITT years= 1	0.01	0.014	-0.021	-0.029	-0.054	-0.418	-0.839	0.307	-0.147
	[0.01]	[0.013]	[0.077]	[0.025]	[0.027]+	[0.455]	[2.206]	[0.410]	[1.004]
ITT years= 2	0.016	0.013	0.161	-0.061	-0.089	1.118	-7.308	0.486	2.127
	[0.011]	[0.014]	[0.107]	[0.041]	[0.053]	[0.633]	[3.090]*	[0.545]	[0.897]*
ITT years= 3	0.015	0.012	-0.071	-0.088	-0.046	1.747	-3.406	0.117	0.439
	[0.01]	[0.016]	[0.073]	[0.035]*	[0.042]	[1.231]	[2.526]	[0.598]	[1.384]
ITT years= 4+	0.014	0.018	-0.07	0.074	0.03	2.634	5.624	-0.101	-1.002
	[0.012]	[0.017]	[0.067]	[0.037]	[0.052]	[1.017]*	[2.953]	[0.537]	[1.935]
School-Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year-Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Level of Observation	student	school	school	school	school	school	school	school	school
Observations	294288	583	531	531	531	583	583	583	583

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
* significant at 5 percent; ** significant at 1 percent..

Table A3

Regression Estimates: Effect of APIP Adoption Duration on the Timing of Freshman Year Entry

	1	2	3	4	5	6	7	8	9
	Freshman within					Sophomore within			
	1 year	2 years	3 years	4 years	5 years	2 years	3 years	4 years	5 years
ITT year=1	0.017	0.026	0.027	0.032	0.034	0.011	0.011	**0.016**	**0.017**
	[0.015]	[0.017]	[0.018]	[0.018]+	[0.018]+	[0.008]	[0.009]	**[0.009]+**	**[0.009]+**
ITT year=2	**0.036**	**0.045**	**0.05**	**0.053**	**0.054**	**0.02**	**0.021**	**0.023**	**0.024**
	[0.016]*	**[0.017]***	**[0.018]****	**[0.018]****	**[0.018]****	**[0.010]+**	**[0.011]+**	**[0.011]***	**[0.011]***
ITT year=3	**0.038**	**0.048**	**0.055**	**0.06**	**0.063**	**0.036**	**0.042**	**0.049**	**0.052**
	[0.019]+	**[0.021]***	**[0.021]***	**[0.021]****	**[0.021]****	**[0.010]****	**[0.012]****	**[0.012]****	**[0.012]****
ITT year=4+	0.017	0.025	0.028	0.033	**0.039**	**0.051**	**0.053**	**0.057**	**0.058**
	[0.024]	[0.024]	[0.024]	[0.024]	**[0.023]+**	**[0.013]****	**[0.016]****	**[0.016]****	**[0.015]****
Observations	224971	224971	224971	224971	224971	224971	224971	224971	224971
R-squared	0.19	0.2	0.21	0.21	0.21	0.13	0.17	0.17	0.17
F: no effect	0	0.01	0	0	0	0	0.01	0	0

Robust standard errors in brackets are adjusted for clustering at the school level.
+ significant at 10 percent; * significant at 5 percent; ** significant at 1 percent.

Table A4

Summary Means and Standard Deviations for Outcomes by Ethnicity

	Black		Hispanic		White	
	Mean	Std.	Mean	Std.	Mean	Std.
Read score 10th grade	-0.257	(1.04)	-0.203	(1.00)	0.247	(0.90)
Math score 10th grade	-0.343	(0.97)	-0.203	(0.95)	0.261	(0.94)
LEP	0.008	(0.09)	0.258	(0.44)	0.006	(0.08)
Low income	0.476	(0.50)	0.574	(0.49)	0.146	(0.35)
Female	0.521	(0.50)	0.505	(0.50)	0.499	(0.50)
Take any AP course	0.225	(0.42)	0.211	(0.41)	0.354	(0.48)
AP courses taken	0.638	(1.52)	0.56	(1.42)	1.186	(2.11)
Take any AP exams	0.034	(0.18)	0.039	(0.19)	0.106	(0.31)
AP exams taken	0.052	(0.33)	0.064	(0.39)	0.203	(0.76)
AP exams passed	0.009	(0.13)	0.021	(0.19)	0.113	(0.54)
Ever a freshman	0.566	(0.69)	0.445	(0.62)	0.773	(0.72)
Ever a sophomore	0.256	(0.52)	0.23	(0.49)	0.445	(0.63)
Ever a junior	0.119	(0.33)	0.095	(0.30)	0.266	(0.45)
Freshman GPA	2.071	(1.20)	2.319	(1.19)	2.599	(1.13)
Graduate with BA	0.09	(0.29)	0.08	(0.27)	0.232	(0.42)
Graduate with AA	0.019	(0.14)	0.029	(0.17)	0.041	(0.20)
Freshman at private college	0.047	(0.21)	0.013	(0.11)	0.067	(0.25)
Freshman at four-year college	0.187	(0.40)	0.099	(0.30)	0.269	(0.45)
Log wage in 2010 (4th quarter)	7.977	(1.37)	8.302	(1.18)	8.37	(1.32)
Wage in 2010 (4th quarter)	5180	(4762)	6252.8	(5081)	7608.67	(7947)
Working in 2010	0.4771	(0.50)	0.4631	(0.50)	0.5229	(0.50)
Observations	69445		128291		83505	

Appendix Table A5: *Effect on AP Exam Subjects Taken*

	Effects on Log AP Exam Taking at School in a Given Year: By Subject				
	1	2	3	4	5
	Math and Computer Science	English	Social Sciences and History	Science	Art and Music
ITT year= 1	0.084	0.285	0.013	0.134	-0.16
	[0.138]	[0.133*	[0.155]	[0.129]	[0.22]
ITT year= 2	0.082	0.403	0.081	0.028	-0.037
	[0.211]	[0.200]*	[0.261]	[0.176]	[0.365]
ITT year= 3	0.294	0.677	0.244	0.441	0.38
	[0.204]	[0.204]**	[0.305]	[0.213]*	[0.423]
ITT year= 4+	0.214	0.804	0.284	0.803	0.289
	[0.25]	[0.238]**	[0.326]	[0.218]**	[0.328]
Observations	570	570	570	570	570

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
* significant at 5 percent; ** significant at 1 percent. All regressions include school- and year-fixed effects.

Appendix Table A6: *Changes in AP Course and Exam Takers after Adoption*

	1	2	3	4	5	6	7	8	9	10
	AP Exam Takers					AP Course Takers				
	School Rank in Tenth-grade Math	School Rank in Tenth-grade Reading	Normalized Tenth-grade Math Score	Normalized Tenth-grade Reading Score	Predicted GPA	School Rank in Tenth-grade Math	School Rank in Tenth-grade Reading	Normalized Tenth-grade Math Score	Normalized Tenth-grade Reading Score	Predicted GPA
ITT years= 1	-4.436	-1.304	0.013	0.02	0.017	-3.029	0.459	-0.015	0.005	-0.008
	[7.667]	[7.233]	[0.022]	[0.014]	[0.010]	[7.558]	[7.099]	[0.022]	[0.014]	[0.007]
ITT years= 2	-14.888	-6.259	0.012	0.002	0.011	-12.418	-7.98	-0.022	-0.003	-0.007
	[9.446]	[8.397]	[0.043]	[0.022]	[0.015]	[10.529]	[9.328]	[0.043]	[0.024]	[0.012]
ITT years= 3	-10.477	-5.166	-0.056	-0.015	0.001	-4.842	-0.799	-0.07	-0.033	-0.012
	[8.673]	[8.10]	[0.05]	[0.022]	[0.017]	[10.27]	[8.994]	[0.051]	[0.026]	[0.014]
ITT years= 4+	-7.994	-5.513	0.053	0.031	0.029	-1.859	-1.139	-0.002	-0.011	0.008
	[8.64]	[8.773]	[0.058]	[0.028]	[0.019]	[12.805]	[11.867]	[0.058]	[0.031]	[0.016]
School FX	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FX	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	18008	18008	18008	18008	18008	61855	61855	61855	61855	61855

Heteroskedasticity robust standard errors in brackets are adjusted for clustering at the school level.
* significant at 5 percent; ** significant at 1 percent.

Appendix Note 1: *Relationships with Donors*

Given that APIP adoption was not random, readers may worry that schools could self-select into the APIP. Because the analytical sample includes only those schools that may have selected into the program and all analyses are based on within-school variation, this type of self-selection will not bias the results, as long as there are no changes in schools that coincide with the exact timing of APIP adoption (which is about two years after when selection would take place). In any case, it is important to ascertain the extent to which schools may have selected into the APIP. One natural question to ask is whether the APIP donors had previous contact with the schools. If so, it would imply that the APIP schools are those types of schools with relationships with donor organizations. However, it does not imply that the timing of adoption is endogenous to changes within schools. Of the schools that were treated in the sample, seven had donors for which there were previous projects.[36] However, none of these schools had any coincident projects that would confound the APIP effects. To ensure that the results are not driven by these schools, I ran the main models excluding these schools, and the results remain unchanged. Of the comparison APIP schools (i.e., those APIP schools that have not yet adopted the program), seven of the Austin schools had relationships with the Michael and Susan Dell Foundation. These seven Austin schools adopted the APIP in 2008, so they serve as comparison schools in my data. However, starting in 2002, these schools all had the Advancement Via Individual Determination (AVID) program, Project Advance, and Project Smart. All of these programs are college-readiness programs that would lead to an *underestimate* of the APIP's effects (because these schools serve as comparison schools rather than as treatment schools in these data). Again, I have determined that the results are robust to excluding these seven schools.

Another important related question is whether any of the donors were involved in other concurrent projects in schools that would confound the APIP effect. While the answer for most school is no, there is one potentially problematic donor relationship that requires some discussion. In five of the Dallas schools that started the program in 2003,[37] the donor offered scholarships to any student who was accepted to college. As such, for these schools, the APIP effect is potentially confounded with a financial aid effect. To ensure that these five schools do not drive the main results, I have replicated the analysis without these five schools, and the treatment effects are slightly *larger* with these school excluded. As such, I can rule out that the few potentially problematic donor relationships bias the results.

Appendix Note 2: *Related Statewide Policies During the Sample Period*

The Texas ten percent rule was put in place in 1997 and ensured that the top ten percent of students from each high school in the state would be guaranteed admission to a Texas public university. One would expect college matriculation rates to have increased in schools that have on average low achievement, such as the selected APIP schools, even if these schools did not adopt the APIP. However, none of the APIP schools adopted the APIP in 1997, so that the timing of adoption is not coincident with the introduction of the new state policy. Furthermore, all the main results are robust to using only those schools that adopted the APIP after 2000.

The Texas statewide Advanced Placement Incentive Program was introduced in academic year 1999-2000. Under this statewide program, the state appropriated $21 million over the years 1998-2000 for the Texas APIP, up from $3 million the previous biennium. The statewide program provides a $30 reduction in exam fees for all public school students who are approved to take AP exams, teacher training grants of up to $450, up to $3,000 in equipment and material grants for AP classes, and financial incentives to the schools of up to $100 for each student who scores 3 or better on any AP exam. One would expect this policy to increase AP participation and effort, even if the selected APIP schools did not in fact adopt the program. However, all the estimated effects exceed those of the statewide program. (**Source**: Texas Education Agency Press Release: "Number of Advanced Placement Exams Taken by Texas Students Increases Dramatically." August 23, 2000).

[36] Dodge Jones Foundation in Abilene (2 in 2003); Perkins Prothro Foundation in Wichita Falls (3 in 2002); Munson Foundation in Denison (1 in 2004); and Fourth Partner Foundation in Tyler (2 in 2002).

[37] Kimball, Roosevelt, Sunset, Thomas Jefferson, and Seagoville High Schools.