

NBER WORKING PAPER SERIES

THE EFFECTS OF AN INCENTIVE-BASED HIGH-SCHOOL INTERVENTION
ON COLLEGE OUTCOMES

C. Kirabo Jackson

Working Paper 15722
<http://www.nber.org/papers/w15722>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2010

This paper integrates material incorporated in Jackson (2010). The author received financial support for this project from the Spencer Foundation. The author thanks Walter Dewar and Gregg Fleisher of AP Strategies, Nina Taylor, Perry Weirich, and Shawn Thomas of the Texas Education Agency, and Susan Brown and Kathy Cox of the Texas Higher Education Coordinating Board. The author acknowledges helpful comments and suggestions from Joshua Angrist, Liz Chesler, Ron Ehrenberg, Roland Fryer, Caroline Hoxby, Clement Jackson, Lawrence Katz, participants at the Dartmouth College Economics Seminar, and the NBER Education meetings. All errors are my own. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by C. Kirabo Jackson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Effects of an Incentive-Based High-School Intervention on College Outcomes

C. Kirabo Jackson

NBER Working Paper No. 15722

February 2010, Revised February 2012

JEL No. I0,I20,I21,J0,J1

ABSTRACT

I analyze the effects of a program that pays both 11th and 12th grade students and teachers for passing scores on Advanced Placement exams on college outcomes. Using a difference-in-differences strategy, I find that affected students of all ethnicities attend college in greater numbers, have improved college GPAs, and are more likely to remain in college beyond their freshman year. Moreover, the program improves college outcomes even for those students who would have enrolled in college without the program. I also find evidence of increased college graduation for black and Hispanic students suggesting that late high-school interventions may confer lasting positive effects on students, and may be effective at improving the educational outcomes of minority students. The finding of enduring benefits when extrinsic motivators are no longer provided is important in light of concerns that incentive-based-interventions may lead to undesirable practices such as “teaching-to-the-test” and cheating.

C. Kirabo Jackson

Northwestern University

School of Education and Social Policy

2040 Sheridan Road

Evanston, IL 60208

and NBER

kirabo-jackson@northwestern.edu

College matriculation and completion rates for low-income and ethnic-minority students in the United States are much lower than those for affluent whites.¹ These disparities are sobering because much of the differences in wages between whites and minorities can be attributed to differences in skills prior to labor market entry (Neal and Johnson 1996).² While there are differences in college going across groups, much of the gaps in educational attainment across sociodemographic groups occur among those who enter college but do not persist (Adelman 1999, Bowen and Bok 1998, Jencks and Phillips 1998).³ Because academic preparedness is key to college success (Tinto 1993, Kalsner 1991) policies that improve scholastic ability of low-income and ethnic-minority students *before college entry* may reduce these differences in college enrollment and college persistence across groups.

Early educational interventions have been found to have large effects on adult outcomes (Currie 2001, Deming 2009) and some argue that that remediation of inadequate early investments is difficult and costly (Cunha and Heckman 2007, Cunha, Heckman and Lochner 2006). However, if academic underperformance is a result of some economic inefficiency (such as imperfect information, student myopia, suboptimal teacher or student effort), late interventions that alleviate such inefficiency could be very cost effective. While there are numerous programs aimed at high school students with the aim of increasing college-going and readiness, rigorous evaluations of these interventions on college outcomes are lacking.⁴ Because students often self-select into college preparation programs, most studies on the efficacy of pre-college interventions on college success are largely descriptive.⁵

In this paper I analyze the effects of a high-school intervention that includes cash incentives for both teachers and students for each passing score earned on an AP exam, teacher training, curricular oversight, and test-prep sessions on students college outcomes. The

¹ Using the August 2006 Current Population Survey, I find that 71 percent of white high-school graduates or GED holders between the ages of 25 and 29 ever enrolled in some college program. The corresponding figures are 60 and 52 percent for blacks and Hispanics respectively. The share of entrants who have earned an AA or BA degree for these same groups are 68% for whites, 51% for blacks and 53% for Hispanics.

² Cameron and Heckman (2001) and Belley and Lochner (2007) find that long-run factors associated with family environment such as parental education account for most of the differences in college-going across ethnic groups.

³ It is well documented that the majority of college attrition occurs in the first year - so that persistence through the first year is a key predictor of subsequent college success (Brawer 1996, Horn 1998, Bradburn 2002).

⁴ For example, evaluations of the GEAR UP program have looked at the effect on college aspiration, but not on actual college outcomes. Source: <http://www.gearupdata.org/GearUpEvaluation.cfm>.

⁵ There are two notable exceptions. Using random assignment, Seftor, Mamun and Schirm (2009) find no effect of Upward Bound overall, but provide some evidence of increased postsecondary enrollment and completion for students with low educational expectations. Also using a randomized design Kemple and Willner (2008) find that career academics improved the labor market outcomes of young men.

Advanced Placement Incentive Program (APIP) is targeted primarily to low-income, minority-majority school districts with a view towards improving college-readiness. New Mexico and New York City have adopted similar programs, while schools in Arkansas, Alabama, Connecticut, Kentucky, Massachusetts, and Virginia have replicated the APIP.⁶ Using school-level data, Jackson (2010) finds that the APIP increases AP participation, improves SAT and ACT performance and may increase college enrollment. The rich student-level data employed in this study allow me to account for student selection, and tests for the underlying mechanisms — providing a more conclusive analysis. Moreover, this study builds on Jackson (2010) by analyzing the effect of the APIP on important longer-run college outcomes.

I investigate how the APIP, which affected 11th and 12 grade students, affected (1) their college attendance (2) sophomore year college persistence (3) college GPA and (4) college completion. I link Texas high-school data to administrative Texas college records — allowing me to compare the college outcomes of students exposed to the APIP to those of students not exposed to the APIP, if they attended college in Texas.⁷ Because the administrators of the APIP did not roll out the program to 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 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 high-school removes (1) the differences in unobserved attributes that make one student take AP courses while another does not and (2) the differences in unobserved attributes that make students from certain schools more likely to excel at college than others. Using changes at similar schools over the same time period as a comparison removes the effects of state policies and influences that may coincide with adoption at some schools.

While the identification strategy removes several sources of bias, there are three remaining endogeneity concerns. The first is that APIP adoption may be endogenous. To address this, I limit the sample to schools that ever adopt the APIP, with similar levels of motivation. The timing of adoption within this sub-sample is determined by the idiosyncratic preferences and

⁶ Lyon (2007), Medina (2007), and Mathews (2004), <http://www.nationalmathandscience.org>

⁷ These data also allow me to account for selection to college, a source of bias in many studies on the pre-college determinants of college success (Breland 1979, Camara and Echternacht 2001).

availability of private donors. Supporting this assertion, I show that outcomes change *after* adoption, conduct falsification tests on other school and principal inputs, and show that the results are robust to including school-specific trends. The second concern is that students may select to schools. I address this in three ways. First, I define “intention to treat” based on enrollment in 10th grade rather than actual enrollment in 11th and 12th grade. Second, I show that the results are robust to eliminating students from non-feeder middle-schools, and making inferences within students who attended *both* the same middle- and high school. Lastly, I show that there is little empirical evidence of selective migration. The last concern is that with Texas college data only, shifting from out-of-state to in-state colleges as a result of the APIP would look like improved outcomes. To address this concern, using National Student Clearinghouse data I show that APIP adoption increases in-state, but not out-of-state, college going. In sum, I am reasonably confident that the estimates presented reflect a real causal effect.

While secondary school interventions may improve contemporaneous student outcomes and increase college-going⁸, there are a variety of reasons why such schemes may not improve outcomes *after* students enroll in college. First, improvements in outcomes may reflect test-taking effort so that the contemporaneous gains in test scores may not persist.⁹ Second, marginal college enrollees may subsequently fail or drop-out if they are not sufficiently college ready. Third, improved AP performance could come at the expense of other important unrewarded skills. 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 introduced.¹⁰ As such, it is important to study the longer-term effects of high-school interventions such as the APIP.

⁸Angrist and Lavy (2010) 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. Angrist, Lang and Oreopoulos (2009) find that cash rewards for academic achievement lead to higher GPAs for females. Berry (2009) finds that performance incentives given to children improve educational outcomes when children have low initial test scores. Also, Dynarski (2008) and Scott-Clayton (2008) study the incentive effects of grade-contingent scholarships among college students.

⁹ There is evidence that test scores can be boosted “artificially” by providing performance incentives on the day of the exam (Braun and Kirsch 2008) or giving students calorie rich meals before an exam (Figlio and Winicki 2005). Also, Glewwe, Ilias and Kremer (2003) find that students did not retain the test score gains associated with teacher cash incentives, while Bettinger (2009) finds a similar lack of persistence over time for student cash rewards. However, Kremer, et al. (2004) find that gains associated with a merit scholarship program for girls in Kenya persisted, and Angrist, Lang and Oreopoulos (2009) find that females offered cash rewards for academic achievement had higher GPAs that persisted after the rewards were provided.

¹⁰ 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).

I find that affected students took and passed more AP exams, and enrolled in college in greater numbers—consistent with Jackson (2010). Moreover, conditional on college enrollment, affected students had higher grades and increased persistence. Improvements were relatively larger for black and Hispanic students and I find suggestive evidence that affected black and Hispanic students were more likely to graduate with a four-year degree. I show that these graduation results are unlikely to reflect faster time to degree. While the short-run outcomes were similar across schools (as in Jackson 2010), the long-run college effects were largest in schools with established AP programs before APIP adoption (and had little AP course expansion) and schools with high-powered incentives — suggesting that increased supply of AP courses was not the driving mechanism, and that increased teacher and student effort are important aspects of the program's success. Also, the program effects grow over time, indicating that the non-incentive aspects of the APIP are important. Guidance counselors credit increased AP participation to increased encouragement from teachers, better student information, and changes in teacher and peer norms — consistent with the APIP reducing suboptimal decisions.

These findings indicate that both maintaining high standards, and increasing participation in rigorous high-school programs can improve college readiness. The results suggest that incentive programs that include resources to turn increased 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 as they show that an inexpensive program targeted to high-school students is effective at increasing educational attainment.

The remainder of this paper is structured as follows. Section II describes the APIP program. Section III presents the theoretical framework. Section IV presents the data. Section V discusses the empirical strategy. Section VI presents the results, specification and robustness tests, shows that the results cannot be driven by changes in out-of-state college going, and presents evidence on mechanisms. Section VII concludes.

II. Description of the AP incentive program

AP courses are common in US high schools and are typically taken by students in 11th and 12th grade. The courses are intended to be “college level” and most colleges allow successful AP exam takers to use them to offset degree requirements.¹¹ The fact that selective colleges pay

¹¹ While this is true in general, some highly selective colleges only allow students to use AP credits to pass out of

attention to a student's AP scores in the admissions process demonstrates that the exams are considered to be revealing about a student's likely preparation for and achievement in college. 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 graded from 1 through 5, where 3 and higher are regarded as a passing grades. AP courses are taught during regular class time and generally substitute for another course in the same subject (AP Chemistry instead of 11th grade science for example), for another elective course, or a free period. While AP courses count towards a student's high school GPA, they are above and beyond what is required for high school graduation. As a rule, AP courses substitute for less demanding activities.¹²

The APIP is run by AP Strategies, a non-profit organization based in Dallas, and is entirely 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. It also includes teacher training 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 teaches students and trains other AP teachers.¹³ Vertical teams also include teachers whose grade precedes those in which AP courses are offered. For example, a vertical team might create a math curriculum starting in 7th grade designed to prepare students for AP calculus in 12th grade. 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 APIP's monetary incentives are intended to encourage participation and induce effort in AP courses. AP teachers receive between \$100 and \$500 for each AP score of 3 or over earned by an 11th or 12th grader enrolled in their course and can receive discretionary bonuses of up to \$1,000 based on results. In addition, lead teachers receive between \$3,000 and \$10,000 annual salary bonus, 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

prerequisites, but not towards regular graduation credit.

¹²**Source:** Executive Vice President AP Strategies and counselors at several Dallas high-schools.

¹³ Jackson and Bruegmann (2009) find that teachers learn from their peers so that vertical teams may be effective.

school, there is variation across schools in the amounts paid. Overall, the APIP can deliver a considerable increase in compensation for teachers.¹⁴

Students in 11th and 12th grade also receive monetary 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 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 examinations during their 11th and 12th grades can earn several hundred dollars. For example, one student earned \$700 in his junior and senior years for passing scores in AP examinations (Mathews 2004). Since students must attend the AP courses *and* pass the AP exams to receive the rewards, students who did not take the AP courses would not take the exams in an attempt to earn the cash rewards. This aspect of the incentives makes them relatively difficult to game and likely to increase overall 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 between 60 and 75 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 for attending team meetings, bonuses to teachers and administrators for AP performance, and some of the supplies and equipment costs. 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.

As a rule, adoption of the APIP works as follows. First, interested schools approach AP Strategies and are put on a list.¹⁵ 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

¹⁴ One AP English teacher in Dallas had 6 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).

¹⁵ There are a few exceptions. Schools in Austin were approached by the donor to adopt the APIP in 2007. Also, five schools in Dallas secured a donor before approaching AP strategies.

within the group of willing schools. In most cases the donor wants a specific district.¹⁶ 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.¹⁷ It takes about two years 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 sciences.

There is variation in the timing of the introduction of the program across schools that I exploit to identify the effect of the program. As illustrated in Figure 1, there are 58 schools that 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.¹⁸ 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. Theoretical Framework

In this section I provide a theoretical framework within which think about how the APIP may affect AP course and exam taking, college going, and subsequent college performance.

III.1 Effect on AP Course and Exam Taking: Student AP output is a function of student and teacher effort in AP courses and exams. Under the APIP, teacher pay is more closely tied to the AP output of their students and the gains to a student of taking and doing well on AP exams are greater. Where good performance on rewarded tasks is more likely with higher teacher and

¹⁶ For example: The first ten Dallas schools were chosen based on proximity to AP strategies; ST Microelectronics is located in the Carrollton-Farmers community and funded this district’s schools; The Priddy Foundation specifically requested the Burk Burnett and City View schools; anonymous donors specifically requested Amarillo and Pflugerville schools; The Dell foundation (headquartered in Austin) funds the Austin and Houston programs; The remaining Dallas schools were funded by the O’Donnell foundation to complete the funding of Dallas ISD.

¹⁷ The seven schools to adopt 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.

¹⁸ For example, in 2005 four high-schools were chosen by The Michael and Susan Dell Foundation from a list of seven willing Houston schools. The remaining three schools may adopt the program at a later date.

student effort, economic agents (in this case, teachers and students) will exert more effort to improve such output (Holmstrom and Milgrom 1991). Therefore, one would expect (A) increases in teacher recruitment efforts to AP courses, (B) increases in teacher effort to improve instruction quality, (C) increases in AP exam taking, (D) increases in student effort on AP exams, and (E) increases in AP course enrollment.

With fully rational students, full information, no supply constraints, and functioning credit markets, the small financial rewards for students of \$100-\$500 for taking AP courses and passing AP exams should have little effect because students will balance the *lifetime* benefits to taking AP courses against the immediate costs. However, the cash incentives may produce large effects if students are myopic, discouraged by their peers and teachers, or credit constrained.

III.2 Effect on College Going: Because the APIP likely increases the number of AP exams students pass, these students may be more desirable candidates and would therefore be more likely to be admitted to college. Also, students can earn scholarships based on their AP performance and can obtain college credit for passing AP scores — reducing the direct costs of college attendance and therefore increasing enrollment. While the APIP should increase college going, conditional on *applying* for college, the APIP could affect students' college *application* decisions for two reasons. First, because the APIP provides information to students about the benefits to taking APs, the program may provide a signal that college is attainable and important such that affected students are more likely to apply. Second, the decision to attend college is an investment under uncertainty. One of the potential benefits of the APIP is to expose students to college-level material, thus providing information to students about the desirability of college and their likelihood of success in college.¹⁹ If students are pessimistic/optimistic about their chances of success at college, the APIP may lead them to adjust their expected costs and benefits of attending college so that they may be more/less likely to apply.²⁰ In sum, the total effect reflects a combination of the effect on college application behaviors and the effect on being

¹⁹ This idea is similar to Costrell (1993), who models the information value of matriculating in college to learn ones suitability. He argues that this could explain the low college completion rates among certain populations.

²⁰ Findings by Stinebrickner and Stinebrickner (2009) suggest the latter is more likely. One implication of this information story is that changes in college application behaviors may reflect an optimal response to new information — such that reductions in the likelihood of applying to college need not be a bad outcome *per se*.

admitted to college conditional on applying. While the APIP should increase the college going of college applicants, the total effect on college going is ambiguous in sign.

III.3 Effect on College Performance: The APIP increases student exposure to more rigorous material either through (1) increased AP participation, (2) increased effort in AP courses, or (3) by improvements in the quality of AP courses. As such, the APIP could have positive effects on college outcomes even with no increase in AP participation. The APIP should increase student knowledge, which should be associated with improved academic outcomes. All else equal, if the APIP only affects students through these three mechanisms, APIP exposure should be associated with improved college outcomes. I refer to this as the human capital mechanism.

However, the APIP may not improve student college outcomes: First, there is evidence that test scores can be improved by increasing test taking effort (Braun and Kirsch 2008), having a good meal (Figlio and Winicki 2005) or gaining familiarity with the test. As such, the increase in AP exams passed may not reflect increased knowledge so that APIP students may not perform any better in college than non-APIP students while in college. Second, a psychology literature suggests that students may be sapped of their intrinsic motivation as a result of being exposed to monetary incentives (Kohn 1999, Deci and Ryan 1985, Cameron and Pierce 1994), so that APIP students may actually perform *worse* in college than students who were never offered rewards. Third, improved AP performance may come at the expense of other important skills if teachers teach-to-the-test or students neglect their non-AP courses. Fourth, the APIP may make students overly ambitious such that they apply to more difficult programs than they would otherwise and have worse outcomes as a result. That is, the APIP could lead to a sort of “mismatch”²¹ between students and colleges that may ultimately lead to worse outcomes.²²

In sum, while a human capital explanation suggests that the APIP would improve college outcomes, the APIP could lead to worse outcomes so that the overall effect of the APIP on long-run college outcomes is an empirical question.

IV. The APIP Schools and The Data

To show how APIP schools differ from other schools in Texas, I present school level

²¹ Under the “mismatch” hypothesis students who wouldn’t ordinarily be admitted to selective schools may be inadequately prepared, so that they would fare better at schools better matched to their preparation (Summers 1970).

²² This mismatch hypothesis is counter to the information story – such that affected students are *more* likely to make sub-optimal college going decisions.

summary statistics from the National Center for Education Statistics and the Texas Education Agency (TEA) in Table 1. Schools that were selected for the APIP were different from schools that have not been selected and may never be 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 a large or mid-sized city 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 for non-APIP schools, and 10 percent of students were limited English proficient at APIP schools compared to less than 4 percent for 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 used combines student records from every public tertiary institution in Texas²³ between 1995 and 2010 and every private institution in Texas between 2003 and 2010²⁴ from the Texas Higher Education Coordinating Board with student-level high-school and middle-school data (including standardized 10th grade test scores that all students must take by state law) from the TEA for the years 1994 through 2007. AP exam data come from the College Board.²⁵ The standardized test scores are normalized and 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).²⁶ The final dataset contains college outcomes, high-school, and middle-school data of all students who were in 10th grade between 1994 and 2007. Using the population of 10th graders allows me to account for attrition that may take place after APIP exposure in 11th and 12th grade.²⁷

I present the pre and post APIP adoption summary statistics for the schools that will have adopted the APIP by 2008 (*note: schools adopt the APIP at different times so that the pre-adoption years differ across schools*). About 22.9 percent of students who were in 10th grade

²³ In Texas there are 145 institutions of higher learning. Of the public institutions, there are 35 universities, 50 community colleges, 9 health related institutions, 4 technical colleges and 3 state colleges. On the private side, there are 39 universities, 2 junior colleges and 3 health related institutions.

²⁴ 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 results are largely the same.

²⁵ TAKS (1994-2003) and TASP (2003-2007).

²⁶ The 10th grade retention rate was about 7% in Texas in 1995, among minorities this figure is over 10 percent. <http://www.tea.state.tx.us/reports/1996cmprpt/04retain.html>

²⁷ For example, if the APIP caused student to drop out of high school in 11th grade before the 11th grade enrolment data are collected, then using the population of 11th graders at will yield results that suffer from attrition bias. Basing all estimates on the population of 10th graders before potential exposure to the APIP avoids such bias.

during the pre-adoption years took an AP course while in high school compared to 30.4 percent in the post adoption period. There were similar increases in AP examination taking where those in 10th grade during the pre-period took 0.097 exams and those in the post period took 0.127 exams during their high school career. The 10th grade math and reading standardized scores were below zero for both periods— indicating that the APIP schools were low achievement schools on average. Also, these scores were slightly lower after adoption than before - suggesting possible negative selection into APIP schools.

The average student before adoption was in 10th grade in 1999 compared to 2003 for the post adoption sample. As such, variables that increase with age such as college attendance and completion are difficult to compare without directly taking age into account (as is done the regression analysis). However, to allow for a simple comparison, I compute enrolment by the time since expected high school graduation. About 35, 46, and 53 percent of 10th graders in the pre-treatment cohorts were a freshman in college within 1, 2, and 3 years of expected high school graduation respectively. The 10th graders in post adoption cohorts were more likely to attend college such that 41, 50, and 56 percent of 10th graders in the post-treatment cohorts were a freshman in college within 1, 2 and 3 years of expected high school graduation, respectively.²⁸ 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.²⁹ Comparing figures for ever being a freshman to the enrolment figures above reveal that 59, 79, and 89 percent of college going occurs within one, two and three years of expected high-school graduation, respectively. Because analyzing enrolment within four years of expected high-school graduation ignores enrolment that occur after four years (one tenth of the variation), I analyze eventual enrolment and control for cohort differences directly.

V. Empirical Strategy

Before presenting the identification strategy, In section V.1, I discuss methodological concerns facing this and similar studies and I present my proposed solutions. In section V.2, I present the identification strategy and discuss the source of plausibly exogenous variation.

²⁸ Texas has the second largest and well developed community college system in the United States so that a sizable share of these students enroll in 2-year colleges.

²⁹ The sophomore variables are not computed for the 2007 cohort, because they would be freshman in 2010 if they enrolled directly after expected high-school graduation. Also, college graduation variables are only computed for cohorts before 2004 because subsequent cohorts would not typically have graduated from a 4-year college by 2010.

V.1 Methodological Issues: Because the APIP affects students in 11th and 12th grade, it affects 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 *do not* control for potentially endogenous covariates such as SAT scores or high-school GPA.³⁰

Other important methodological issues are the choice of population and how the outcomes are defined. If the APIP increases college going, the population of college students who were exposed to the APIP may have worse outcomes merely because APIP-exposed students will have a higher proportion of marginal (and likely less well-prepared) college attendees. Using the sample of college *attendees* leads to sample selection bias. To avoid this, I compare the outcomes of *all potential* college students. Basing the estimates on prospective college students allows one to uncover the true causal effect, but introduces the methodological issue that changes in outcomes such as sophomore year persistence and GPA may reflect effects through increased college enrolment and effects due to changes for students who would have enrolled in college absent the APIP (effects conditional on college entry).

While it is impossible to identify those who would have attended college absent the APIP I use three approaches to estimate of the effect of the APIP conditional on college entry. The first approach is to impute outcome values for students who do not enroll in college. Specifically, I predict freshman year GPA and sophomore year enrollment as a function of observable pre-treatment characteristics and use the predicted values for those who do not enroll in college. The second approach is a control function methodology. Specifically, following Angrist (1995) I use the sample of college enrollees, while controlling for the estimated likelihood of attending college (estimated using the full sample).³¹ The third approach is to trim the data following Lee (2002) to obtain a lower-bound estimate of the effect for those who would have enrolled in college absent the APIP.³²

The final methodological issue is how treatment is defined. Because students may enroll

³⁰ Many studies mistakenly control for endogenous variables to isolate the causal effect of AP exams are (Geiser and Santelices 2004, Eimers 2003). However, Dougherty, Mellor and Jian (2006) control for 8th grade test scores.

³¹ I estimate a probit model of attending college on all control variables and obtain an estimated propensity.

³² One estimates the proportion X of the treated group that attends college because of the APIP. Under the extreme assumption that the marginal enrollees have the best outcomes, one ranks the treated sample by the outcome and removes the top X percent of the treated sample. Estimates using the trimmed sample of college enrollees will yield a lower-bound estimate of the effect conditional on enrolling.

at APIP schools in 11th and 12th grade to benefit from the program, defining treatment based on actual school enrollment in 11th and 12th grade could be subject to self-selection bias. To avoid such bias, I use intention-to-treat instead of whether a student is actually affected by the APIP. Specifically, I define intention-to-treatment (ITT) based on whether a student would be treated if they remain in their 10th grade high school and are never held back a grade. For example, a student is intended for treatment if they are enrolled at a school in 10th grade in year t , and the school will have adopted the APIP by year $t+2$. The benefit of using ITT is that it is not endogenously determined by student selection into APIP schools in 11th and 12th grade, or subject to biases due to attrition or retention.³³ Using ITT yields a clean policy-relevant estimate of the effect of introducing the APIP on the target population.

V.2 Identification Strategy: The identification strategy is to compare the difference in college outcomes across cohorts of students who attended the same high-school before and after APIP adoption to the difference in college 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 college outcomes of students with the same 10th 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.³⁴

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 had 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

³³ 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 10th grade will not be treated but will be intended for treatment, (2) students who enter APIP schools after 10th 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.

³⁴ For a description of such policies see the Appendix Note 2.

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.³⁵

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 where donors had prior relationships and verify that all the main results are robust to excluding these schools.³⁶ Also, in section VI, I show that improvements only take place *after* APIP adoption, and the timing of adoption is unrelated to other school changes or the timing of the arrival of a new principal - suggesting the assumption of exogenous timing of adoption is valid.

This within-school cohort-based comparison is implemented by estimating the following equation by Ordinary Least Squares (OLS).

$$[1] \quad Y_{ich} = \beta_1 X_i + \beta_2 A_i + \sum_{k=1}^{4+} \mu_k I_{ITT \text{ year}=k} + \theta_h + \theta_c + \varepsilon_{ich}$$

In [1], Y_{ich} is the outcome of student i in 10th 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 10th grade. A_i is a vector of student achievement scores from 10th 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 $I_{ITT \text{ year}=k}$ 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 k^{th}

³⁵ Some schools adopt 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.

³⁶ For a detailed discussion see Appendix Note 1.

intention to treat year. Standard errors are adjusted for clustering at the school level.

It is important to point out that a simple before/after comparison likely understates the full effect of the APIP because the first affected cohort (ITT year=1) 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, I estimate dynamic treatment effects in addition to a simple before and after comparison. To identify the dynamic APIP effect, I use four binary variables denoting the first, second, third, and fourth plus intention-to-treatment years. For example, the first ITT cohort for a school has ITT year=1 and the 3rd ITT cohort for a school has ITT year=3, so that ITT year denoted how long the APIP had been in place when the student was expected to graduate from high school. For example, if the APIP was adopted in school h in the 2002-03 school year, the 10th grade cohort for the school year 2000-01 would be coded as ITT year=1, while the 10th grade cohort for the school year 2002-03 would be coded as ITT year=3).

VI. Results

VI.1 Graphical Evidence: Before showing the regression results, it is helpful to present some visual evidence. 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 -5 through 4 (the first year of adoption is year 0). For AP course taking, AP exam passing, and freshman year enrollment, there is a clear increase after APIP adoption. However, there is some visual evidence of upward trends in AP course taking and college enrollment (I test for trending formally below). For the longer-run outcomes, freshman year GPA, sophomore year enrollment, and sophomore year persistence (being a sophomore conditional on ever being freshman) there is clear visual evidence of improved outcomes after APIP adoption and little evidence of pre-existing trends. The improved outcomes conditional on college entry are evidence irrespective of how one accounts for sample selection.

To test for trending formally, I test the null hypothesis that the pre-adoption year effects differ from the first pre-adoption year, and I fail to reject at the 5 percent level *for all outcomes*. In contrast, *for all outcomes*, the test that the post-adoption years differ from the year prior to adoption is rejected at the 1 percent level — statistical evidence of an APIP effect (including

conditional on college enrollment) that is not driven by underlying trends. Moreover, in section VI.4, I show that the main findings are robust to including school-specific linear trends.

VI.2 Effects on AP outcomes and college enrollment: 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) a there is a statistically significant 0.165 increase (about a 21 percent increase) in year three, but a statistically insignificant 7.4 percent increase in year four and beyond. Even though the simple before/after coefficient is not statistically significant, the null hypothesis that all the dynamic APIP effect are zero is rejected at the 5 percent level — indicating that the APIP does increase AP course taking. Columns 2 and 3 show that the increases AP exam taking and passing and that this effect increases over time. 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% increase) in the number of AP exams taken, and a 0.043 increase (about a 45% increase) in the number of AP exams passed. The increases are primarily in the English and Science AP exams (Appendix Table 4). It is important to point out that the APIP may affect unmeasured outcomes (such as aspirations or self-confidence) that could affect 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".

Columns 4,5, and 6 show the effect on the likelihood of ever being a freshman in any college, a two-year college, and a four-year college, respectively. The before/after comparisons show that the APIP increases college going by about 4.2 percentage points on average and that this is driven by both two-year and four-year college going. The dynamic effects show that the effect of the APIP increases over time, so that by the fourth year of the APIP college going overall increases by 4.8 percentage points (an 8 percent increase), four year college going increases by 2.8 percentage points (a 16 percent increase) and two year college going by a statistically insignificant 2 percentage points (a 5 percent increase). Also, there were small

statistically insignificant increases in college selectivity among college goers (not shown).³⁷

Because the outcomes analyzed are using Texas college data, it is important to ensure that increased college enrolment does not reflect shifting of college going from out-of-state to in-state. I test this directly using data from the National Student Clearinghouse (NSC) for the 10th grade cohorts of 2005 and 2006. These data measure college enrollment both at in-state and out-of-state colleges and universities. 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 7 and 8 show that by the second ITT year the APIP increases in-state college going by 3.5 percentage points (significant at the 5 percent level) but has no effect on out-of-state college going (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 going numbers from the THECB data and showing that the enrollment effects measured in this study reflect real increases in college going and are not confounded with changes in out-of state college going.

IV.3 Effects on Freshman GPA: In columns 1 through 3 of table 4, I present the effect on Freshman year GPA *conditional on college enrollment*. In models of college enrollees that include the estimated propensity scores to account for selection to college (column 3), treated cohorts had freshman year GPAs that were 0.03 points higher (significant at the 10 percent level). This GPA effect increases over time so that GPAs were 0.067 points higher in the fourth treated cohort than in untreated cohorts (significant at the 5 percent level). Using all students and imputing GPAs for non-enrollees (based on observable characteristics) yields a before/after effect of 0.015 points (significant at the 5 percent level), and a fourth year effect of 0.031 grade points (significant at the 1 percent level). I also implemented a trimming procedure to obtain a lower bound effect. However, the *lower-bound* effect is large negative, and therefore uninformative and not presented in the table. To shed further light on how much of the APIP effect is driven by the extensive margin (college entry) as opposed to the intensive margin (GPA conditional on enrolling in college), in Figure 3, I present the coefficient on the simple post adoption indicator variable using the full sample under different imputed GPA values for those who do not attend college. If one assumes students who do not enroll in college would have

³⁷ I linked the college identifiers for college enrollees to the mean SAT scores of admitted students from IPEDS. Treated cohort attended colleges with slightly higher SAT scores. The increase in the 25th percentile of combined math and reading SAT scores was 32 points (not statistically significant).

received a GPAs of 0, 1, and 2 (an F, D, and C average, respectively) then the before/after effects would be 0.048, 0.034, and 0.021 grade points, respectively. If one makes the unrealistic assumption that non-enrollees would have had the same GPA as those who did enroll (2.3 about a C+ average) the adoption effect would be 0.017 grade points (significant at the 5 percent level). In fact, the 95 percent confidence interval of the estimated APIP effect to includes zero for all assumed values below 2.7 (a B- average), the point estimates are positive until an assumed GPA of 3.6 (an A- average), and is essentially zero (-0.006) if one assumes that non-enrollees would have had a 4.0 (straight As). Figure 3 shows that under reasonable assumptions of what the GPAs would have been for non-enrollees, APIP exposure is associated with improved freshman year GPAs conditional on college enrollment.

To provide a sense of where in the achievement distribution these improvements in freshman GPA might come from, Figure 4 presents cumulative densities of the freshman GPA of treated and untreated cohorts. It is evident that there are slight improvements in GPA for treated cohorts below the 40th percentile (GPA between 0 and 2), but that the 40th through 90th percentiles are clearly higher in the treated group (corresponding to GPAs between 2 and 4). This suggests that much of the improved GPAs came from student between the 40th and 90th percentile of the achievement distribution in college who would have had averages between a C and an A- without the APIP.

IV.4 *Effect on sophomore year enrollment and persistence:* Most college attrition occurs in the first year so that persistence through the first year is a key predictor of college success. In Table 4 I present the effect on being a college sophomore and persisting to sophomore year (enrolled as a college sophomore conditional on ever being a college freshman). Column 4 shows the results for the full sample, and columns 5,6,7 and 8 show the results conditional on college enrollment with no sample selection correction, with propensity scores, with imputed values for non-enrollees, and the lower-bound trimming effect, respectively.

APIP adoption is associated with a 3 percentage point increase in ever being a college sophomore on average, and a 6 percentage point increase for the fourth ITT cohort and beyond. All estimates of the effect conditional on college enrolment yield pooled adoption effects of about 2.5 percentage points and fourth year effects between 5.5 and 8.6 percentage points (all significant at the 1 percent level). Given a base persistence rate of 52 percent, these increases represent increases between a 10 and 17 percent. The results that trim to top GPA earners in the

treated sample for a lower-bound effect yield a statistically insignificant pooled effect of 1 percentage point, but a 6.6 percentage point increase by year four (significant at the 5 percent level). These results show real increases in sophomore year attendance, due, in large part, to increased persistence for those who would have enrolled in college absent the APIP.

Another way to obtain a lower-bound persistence effect is to make the extreme assumption that all the marginal freshman enrollees would persist to sophomore year (Dynarski 2008). The average persistence rate is 0.52 so this is unlikely. The fourth year freshman and sophomore enrollment effects are 4.8 and 6 percentage points, respectively. If all the marginal enrollees would have persisted, this leaves $6-4.8 = 1.2$ percentage points that must be due to the non-marginal students. As such, even under the lower-bound assumption, the APIP increases college persistence conditional on college enrolment by *at least* 2.3 percent by year four. A more realistic, *but still conservative*, assumption is that the marginal student would persist at the same rate as the average student. Under this assumption $4.8*0.52 = 2.5$ percent may be due to marginal enrollees, so that $6-2.5=3.5$ percentage points is due to the infra-marginal students (yielding a 6.7 percent increase in sophomore year persistence conditional on enrolling).

Timing of the Enrollment effect: Readers may wonder if these enrollment effect merely reflect that the APIP causes students to enroll in school sooner. To assess this, I analyze enrollment within 1,2,3 and 4 years of expected high school graduation in Table 5. 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, one would see effects on "ever a freshman within 1 year of high-school graduation" but would see no effect on "ever a freshman within 4 years of high-school graduation". The effects on freshman enrollment and sophomore enrollment grow as one looks to longer time horizons, which is exactly what one would see if the effect reflects increased college going overall, and is inconsistent with the results being due to shifting to earlier college entry. This also indicates that much of the increased college going occurs after four years after expected high school graduation, so that analyzing short run college going would miss an important part of the story.

VI.5 Threats to Validity and Endogeneity Concerns: 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, there are a few remaining endogeneity concerns regarding whether the estimates reflect a true causal relationship. I address these in this section.

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. There is no visually evidence of this and I fail to reject the null hypothesis of improvements in outcomes *before* adoption. As a direct test of the "timing of interest" hypothesis, I regressed the outcomes on the two-year lag of adoption. This yielded very small coefficient estimates and *p*-values larger than 0.4 for all outcomes.

To provide further evidence that the timing of APIP adoption is exogenous, I predict having a new high school principal as a function of whether program will be adopted in 3 years, 2 years, 1 year, or was adopted in the same year. These models include school fixed effects and year fixed effects only. In each of these 4 regressions (shown in appendix Table 4) the *p*-values associated with the null hypothesis of no systematic relationship is larger than 0.2. I also estimate a specification similar to equation (1) and find no effect on subsequent principal turnover (Appendix Table 6). This is consistent with assertions that timing of adoption is idiosyncratic, and suggests that adoption is likely exogenous to *changes* in schools over time.

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.³⁸ If there were positive selection driving the results, the APIP should be associated with characteristics associated with better outcomes. To test for this I predict the main outcomes as a function of observable student characteristics *before* APIP

³⁸ 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 school, or another school in the district (including magnet and charter schools) subject to space limitations at the receiving school. Because Houston and Austin schools do 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 school are the lower performing schools.

adoption. 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 have an effect on these predicted outcomes. For all the predicted outcomes the point estimates are close to zero and the *p*-value associated with the hypothesis that the adoption year variables are correlated with the predicted outcomes are above 0.4 (Table 6).³⁹ To assuage any concerns that the results are driven by controlling for selection I show that results that exclude student controls (Appendix Table 1) are similar to those in Table 3.

There may be selection in unobserved dimensions. To ensure that 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 that 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 non-feeder middle schools to those of students who attended the natural feeder middle schools and did not self-select by making inferences based on the within middle-school-by-high-school variation. That is, 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 (Appendix Table 1) are similar to the results in Table 3— again, suggesting no selection.

APIP schools were already on a trajectory of improvement before adoption. The visual evidence shows and the test for pre-existing trends indicate that the results are not driven by underlying trends. However, it is instructive to see that the results are robust to including high-school trends. I augment the main estimation model to include both a high-school specific intercept and a linear time trend for each high school (Appendix Table 1). While less precise, the estimates are similar to those in Table 3 with the exception of AP course taking, where the estimated effect is *larger* with the inclusion of a linear time trend for each high-school.

The benefits of the APIP are driven by general improvements in schools. One may wonder if

³⁹ I also present effects on individual covariates in columns 1 through 6 which indicate that treated cohorts had lower 10th grade test scores, were less likely to be low income and less likely to be Hispanic. While there are these differences, overall treated students had very similar predicted outcomes as untreated students.

the benefits to the APIP are driven by improvements in or increased effort in the AP program, or due to other school-wide changes (such as better inputs, or change in teaching philosophy) that might have been affected by APIP adoption. If improved outcomes are driven by improved AP instruction, increased AP participation, or both, one should see large effects for students who take AP courses and little effect for students who do not. I test this by comparing the effect of the APIP on students who are *ex-ante* (based on pre-adoption characteristics) likely and unlikely to take AP courses (Figure 5).⁴⁰ While one cannot reject the null hypothesis of no adoption effect for those with estimated likelihood below 0.33 at the 10 percent level, there are large statistically significant APIP effects for students with estimated likelihood above 0.66. This suggests that (a) the benefits of the APIP are experienced by AP students and (b) the estimated effects are *not* driven by other confounding changes at schools.

As an additional test, I look at the APIP effect on school-level outcomes that should not be affected by the APIP as a falsification exercise. Appendix Table 5 shows the effect of APIP adoption on high-school graduation, teacher turnover, total school expenditures, the number of teachers, teacher experience, and average class size. By the fourth year of APIP adoption the only outcome for which there is a statistically significant effect is the number of AP teachers - consistent with an expansion of the AP program. For all other outcomes, there is no significant effect, suggesting that the estimated effects are not due to general improvements in schools.

VI.5 Effects by Gender and Ethnicity. In light of a literature showing larger positive treatment effects for females than for males (Kling, Liebman and Katz 2007, Jackson 2009, Hastings, Kane and Staiger 2006, Angrist, Lang and Oreopoulos 2009, Angrist and Lavy forthcoming), one wonders if there is response heterogeneity by gender. Also, Klopfenstein (2004) documents large differences in AP participation across ethnic groups both across *and within schools* so that there may be differences by ethnicity. As such, I estimate equation [1] with the inclusion of interactions of the ITT year indicator variables with "female", "black", "Hispanic", and "other-ethnicity" indicator variables. I report the adoption variables and their interactions in Table 8.

In general the estimated effects are larger for females than for males, however these differences are not large. Of the 10 outcomes analyzed, only for two of them does one reject the

⁴⁰ I estimate a probit model of the likelihood that a student takes at least one AP course based on all observed pre-treatment covariates. I then put those with estimated propensities higher than 0.66 into the highly likely group and those with estimated propensities below 0.33 in the low propensity group.

null hypothesis that there are gender differences in response at the 10 percent level. Specifically, females have a smaller increase in AP courses taken and about a 50 percent larger increase in overall college going by the fourth APIP year. For all outcomes conditional of college entry, the result are similar for males and females, such that while females were more likely to enroll in college due to the APIP, the effect on their college outcomes are similar.

Before analyzing differences by ethnicity, I present summary statistics in Table 7 for white, black, and Hispanic students. About 23 percent of black 10th graders in the sample take any AP course, only 3.4 percent take any AP exams and the average number of exams passed is 0.009. Similar to black students, about 21 percent of Hispanic 10th graders in the sample take any AP course, only 3.9 percent take any AP exams and the average number of exams passed is 0.021. In contrast, about 35 percent of white 10th graders take any AP course, 10.6 percent take any AP exams, and the average number of exams passed is 0.113. Not surprisingly, these differences in AP outcomes are associated with differences in college outcomes. The college going rates for blacks, Hispanics, and whites are 56.6, 44.5, and 77.3 percent, respectively. Even conditional on enrolling, there are differences so the freshman year GPAs for blacks, Hispanics, and whites were 2.07, 2.32, and 2.6, respectively. It is worth noting that some of these differences may reflect differences in preparation as black and Hispanic 10th graders had math and reading test scores between 0.4 and 0.5 standard deviations lower than whites on average.

The results in Table 8 reveal differences across ethnic groups. While there are small differences in AP course taking and AP exam taking, one rejects that the APIP response is the same between white, black and Hispanic students at the 5 percent level for college going, GPA and persistence to sophomore year. Given the large differences in baseline AP outcomes, the increases in AP exam taking represent a near doubling of AP exam taking for whites an increase by a factor of 2.5 for Hispanics and a near tripling for blacks. The relative differences for AP exam passing are even more dramatic; a 34 percent increase for whites, a doubling for Hispanics and a five times increase for black students.

The differences in college outcomes across ethnic groups reveal some interesting patterns. While black and white students experience increases in college going, there is no increase in college-going for Hispanics. Looking to GPA conditional on enrolling, while white students have little response in this outcome, by the fourth APIP year, black students have freshman year GPAs that are between 0.138 and 0.161 higher than those not exposed (these effect are significant at the 1 percent level). While not statistically significant, the point estimates suggest that Hispanic students may also have higher GPAs under the APIP. All groups have improved sophomore year enrolment and persistence conditional on enrollment, and the improvements are particularly pronounced for black students. In models that use the propensity score correction, by the fourth APIP cohort, sophomore year persistence increases by 4.85 percentage points for whites, 9.8 points for Hispanics, and 13.5 points for blacks (these differences are statistically significant at the 1 percent level). However, in models that use imputation to account for sample selection bias, by the fourth APIP cohort, sophomore year persistence increases by 4.85 percentage points for whites and Hispanics, and 9 points for blacks (this differences are statistically significant at the 1 percent level). In sum, while there are small differences by gender the APIP appears to have relatively larger effects on the college outcomes of ethnic minorities, particularly black students. The pattern of results implies that the APIP may help narrow educational gaps across ethnic groups.⁴¹

VI.6 Evidence on College Graduation: Analyzing college completion with these data is feasible, but results must be interpreted with caution. Most APIP schools adopted the program after 2000, so that those who would graduate after 2010 are classified as non-graduates in these data. To reduce the likelihood of classifying people still in school as non-graduates, I analyze

⁴¹ Given the differences in academic preparation across the ethnic groups, readers may wonder if these differences by ethnicity merely reflect differences by incoming academic preparation. To test for this, I estimated the main models that interacts indicator variables denoting whether a student scored in the top third or bottom third in both math and reading 10th grade tests with the adoption variables. There were no systematic or statistically significant differences by incoming academic preparation, and they do not explain the differences observed by ethnicity. This suggests that social factors were at play.

college graduation for the students who were in 10th grade before 2003. These students are old enough to have graduated from a 4-year college by 2010. An implication of this data truncation problem is that I may confound reduced time to degree with increased college completion. However, I present evidence that this is likely not the case so that this truncation bias is likely to attenuate the estimates and *understate* the effects of the APIP on graduation.

Columns 1 and 2 of Table 9 presents the main specification where the outcomes are graduating with any degree, graduating with any degree within 4 years, and graduating with a bachelors (BA) degree. I present results for all students in the top panel, and also black, Hispanic, and white, separately in the lower panels. The results in the top panel show that the APIP has no effect on the graduation variables, on average. However, the results broken up by ethnicity reveal that while there are no graduation effects for whites, treated black and Hispanic students were more likely to earn a bachelors degree (one can reject the null hypothesis of no adoption effect at the 5 percent level for both groups). For black students, the first and third APIP cohorts are 1.8 percentage points more likely to earn a BA degree than those who were not exposed to the APIP, while the effect for Hispanic students increases monotonically with program age. By the fourth year of the APIP Hispanic students are 2.5 percentage points more likely to earn a BA degree. These increases represent a 20 and 31 percent increase for black and Hispanic students, respectively. The large effects for Hispanics are striking given the lack of a college enrollment effect— indicating that this effect is all conditional on college entry.

To assess whether these effects reflect decreased time to degree rather than increased degree attainment overall, I analyze the effect of earning a degree within 4 years of expected high school graduation. As with the enrollment effects, if the effects were due to black and Hispanic students earning degrees faster, one would see stronger effects on this variable that only measures degree receipt close to high school graduation. That is, the effects would increase as we look at degree receipt within shorter time horizons and disappear as one looks to long time horizons. There is no effect on earning a degree within four years for any group — showing that the degree effect is not driven by students being more likely to graduate on time. In fact one has to look seven years out to find effects on graduation similar to those for the ever earned a BA degree variable. To provide further evidence that this graduation effect reflects increased persistence and more completed schooling (rather than shorter time to degree), I look at total credits earned and total credits earned within four years of expected high school graduation.

Again, there are larger positive effects on total credits earned than on credits earned within four years of expected high school graduation — further evidence that the effects are due to increased schooling rather than reduced time to degree. As a final check, I look at junior year enrollment and junior year enrollment within four years of expected high school graduation. The pattern is the same; all groups have modest increases in junior year enrollment within four years and larger effects on ever enrolling as a junior — further evidence that the effect are due to increased schooling rather than reduced time to degree. To assuage any lingering concerns that these findings are due to differential trends by ethnicity, I estimated models with school-specific trends by ethnicity and the results are qualitatively similar.⁴²

VI.8 Evidence of the Mechanisms: Having established that there is an effect of the APIP, in this section, I try to shed light on the underlying mechanisms with a variety of empirical tests.

Are the improvements driven by increased supply of AP courses and sections? It is natural to wonder whether the improved outcomes were driven by merely an increase in the availability of AP courses and sections. To speak to 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 those that had below the median (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 in the high growth schools and small effects in the low growth schools. Exactly the opposite is true. There are small improvements in college outcomes for the high growth schools and large improvements for the low growth schools (Figure 5) — indicating that increases on AP course supply do not drive the results.

Are schools/teachers/students motivated by the incentives? If better resource utilization (caused by the monetary incentives) drives the success of the APIP, and effort is proportional to the size of the rewards, schools that paid higher powered incentives would have better outcomes than those that did not. To test this, I compare the adoption effect for high power schools (paid between \$101 and \$500 per exam) and low powered schools (paid \$100 per exam) in Figure 5. While the effect on AP exam passing is similar across these school types (as found in Jackson

⁴² For visual evidence Appendix Figure A1 shows the evolution of the graduation and junior year enrollment for black white and Hispanic students. The F-tests on pre treatment years yield p -values 0.21, 0.36, and 0.09 for white, Hispanic and black students respectively, suggesting some possible trending *only for black students*.

2010), for all college outcomes, the effects are larger in the high-power schools. This suggests that the monetary incentives, and the increased effort they induced, were an important component of the success of the program.

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 are driven by incentives alone, the effects would be the same in the second year as all subsequent years. For most outcomes this is not the case, suggesting that there is learning by doing or the other components of the APIP that 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 looking at the incoming 10th grade test scores. The results in Table 4 show that incoming test scores are slightly *lower* after adoption than before, suggesting that improvements in incoming academic preparedness is not the driving force. This suggests that learning-by-doing and changes in school culture are the likely explanation for the increases in the APIP effect over time.

IV.9 Discussion: The large increases in AP participation are difficult to reconcile with the full-information full-rationality models of the schooling decision, and are more consistent with there being sub-optimal effort on the part of teachers and/or students or a large increase in the supply of AP courses. Guidance counselors at three different APIP high schools in Dallas indicate that there were school-wide campaigns to increase participation in AP courses after APIP adoption. At two of the three high schools an additional guidance counselor was hired to improve the school's 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 sold 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, which could increase the student's likelihood of being in the class's top ten percent and gaining admittance into a good college. Guidance counselors mentioned a shift in student and teacher attitudes toward AP courses such that AP

courses are no longer considered only for the very brightest of students.

The tests above suggest that increased supply of courses was not the driving force behind the AP participation response or the improvements in college. The fact that the effects are larger in the 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 story. The fact that the effects grow over time suggest that the program aspect emphasized by guidance counselors such as information and outreach (that schools will have gotten better at over time) are also important. The body of evidence indicates that all aspects of the program are important and that providing cash incentives to students or teacher alone, or providing teacher training alone, or expanding the AP course offerings alone, would not have yielded the same results as the full intervention.

VII Conclusions

Using a carefully selected 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 (with no effect on out-of state college going), had higher GPAs, and were more likely to persist beyond their freshman year. These improvements were due both to increased college enrolment and improvements in college outcomes for those students who would have enrolled in college absent the program. 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, other coincident school policies.

While there small differences by gender, the APIP led to larger improvements in student performance and persistence conditional on college attendance for black and Hispanic students than white students. Looking to college graduation, there is little evidence of an increase overall. However, there is suggestive evidence of increased college graduation for Hispanic and black students. This implies that programs like the APIP may be effective at reducing some of the educational gaps that persist across ethnic groups.

Given that I find no evidence of worse outcomes associated with the APIP, these improvements in college outcomes were likely the result of increased exposure to rigorous material induced by the APIP. Consistent with this interpretation, APIP adoption is associated with increased AP course taking, AP examination taking, and the effects only exist for students

who were *ex ante* likely to take AP courses. The evidence on mechanisms indicate that both the incentive aspects and the non-incentive aspects are important. The finding that the program confers enduring benefits on students when extrinsic motivators are no longer provided is important for the literature on student and teacher incentives in light of concerns that incentive-based-interventions may lead to undesirable practices such as “teaching-to-the-test” and cheating. More generally, 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 recourses to help translate increased effort into results.

To get a sense of the cost-effectiveness of the APIP, consider the following conservative back-of-the-envelope cost/benefit calculation. The program costs about \$200 per student who takes an AP exam per year. Roughly 7 percent of 10th graders take an AP exam after APIP adoption so the cost per 10th grader is about \$14. Assuming a student is affected for two years, this comes to \$28 per 10th grader. By the fourth year of implementation, the APIP increases the likelihood of ever being a sophomore by 6 percentage points. Under the conservative assumption that those students only complete one additional year of school, this would lead to an increase in the average overall years of schooling of 0.06 years. This implies that using the APIP program, one can increase the average years of educational attainment by 1 year at a cost of about \$466 per 10th grade student. For this program to not be cost effective would require the implausible scenario that the present discounted value of an additional year of education is less than \$466. As such, this program is likely a worthwhile investment.⁴³ Given that the large increases in AP participation imply that low AP participation may reflect some sub-optimality, and anecdotal evidence from guidance counselors that the increased AP participation was the result of increased information, changes in peer norms, and reduced barriers to taking AP exams, it is not surprising that the economic returns to the program are large. The large effects of the APIP imply that it may be possible improve the outcomes of students by improving their decision making and increasing access to well taught rigorous courses.

⁴³ To make this point more clear consider the following calculation; suppose the rate of return to an additional year of education was one percent. For someone earning the median household income of approximately \$42,000 per year a one percent increase in wages would provide an additional \$420 per year. For a worker with 35 years of work ahead of them, at an interest rate of 10 percent, an additional \$420 per year is worth a lump sum payment of \$4035 today. This is so much larger than the per pupil cost of \$300 that the rate of return to education would have to be much less than one percent for this program not to be cost effective.

Overall, the findings suggest that providing monetary incentives to both students and teachers to promote increased participation and improved performance in rigorous courses in addition to providing additional resources to support the increased efforts can lead to meaningfully improved student outcomes. The fact that the positive effects were larger for ethnic minority students suggests that similar program programs may help reduce some of the educational differentials that currently exists across ethnic and socioeconomic groups. In light of research on the efficacy of early versus late interventions, these findings are noteworthy because they suggest that a relatively inexpensive program targeted relatively late in a student's educational career can increase their eventual educational attainment to a considerable degree and likely has a high rate of return.

Bibliography

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

Angrist, Joshua. "Conditioning on the Probability of Selection to Control Selection Bias." *NBER Technical Working Paper 181*, 1995.

Angrist, Joshua, and Victor Lavy. "The Effects of High Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial." *American Economic Review*, forthcoming.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 1,1 (2009): 136-163.

Angrist, Joshua, Eric Bettinger, and Micheal Kremer. "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review*, June 2006.

Belle, Philippe, and Lance Lochner. "The Changing Role of Family Income and Ability in Determining Educational Achievement." *Journal of Human Capital* 1, no. 1 (2007): 37-89.

Berry, James. "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India." *unpublished mimeo Cornell University*, 2009.

Bettinger, Eric P. "Paying to Learn: The Effect of Financial Incentives on Elementary Test Scores." *unpublished* , 2009.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *NBER Working Paper 15361*, 2009.

Bowen, William, and Derek Bok. *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press, 1998.

Bradburn, Ellen M. *Short-term Enrollment in Postsecondary Education: Student Background and Institutional Differences in Reasons for Early Departure*. Washington, D.C.: U.S. Department of Education National Center for Education Statistics, 2002.

Braun, Henry, and Irwin Kirsch. "Testing the Effect of Incentives on Student Performance on the NAEP." *unpublished mimeo*, 2008.

Brawer, F.B. "Retention-attrition in the nineties." *ERIC Document Reproduction Service No.*, 1996.

Breland, Hunter M. *Population validity and college entrance measures*. New York: College Board, 1979.

Camara, W. J., and G. Echternacht. *The SAT I and high school grades: Utility in predicting success in college*. Research Notes, The College Board, Office of Research and Development., 2001.

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

Cameron, Stephen, and James Heckman. " The Dynamic Models of Schooling Attainment for Blacks, Whites and

Hispanics." *Journal of Political Economy*, 2001.

Costrell, Robert M. "An Economic Analysis of College Admission Standards." *Education Economics* 1, no. 3 (1993): 227-241.

Cunha, Flavio, and James Heckman. "The Technology of Skill Formation." *American Economic Review* 97, no. 2 (2007): 31-47.

Cunha, Flavio, James J. Heckman, and Lance, Lochner. *Interpreting the Evidence on Life Cycle Skill Formation*. Elsevier, 2006.

Currie, Janet. "Early Childhood Education Programs." *Journal of Economic Perspectives* 15, no. 2 (2001): 213-238.

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

Deming. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1, no. 3 (2009): 111-134.

Dougherty, Chrys, Lynn Mellor, and Shuling Jian. *The Relationship between Advanced Placement and College Graduation*. AP Study Series Report 1, February, National Centre for Education Assessment, 2006.

Dynarski, Susan. "Building the Stock of College-Educated Labor." *Journal of Human Resources* 43, no. 3 (2008): 676-610.

Eimers, Mardy. "Dual Credit and Advanced Placement: Do They Help Prepare Students for Success in College?" *MidAIR and AIR Conference*. Tampa, Florida, 2003.

Figlio, David N., and Joshua Winicki. "Food For Thought: The Effects Of School Accountability Plans On School Nutrition." *Journal of Public Economics* 89 (February 2005): 381-394.

Figlio, David N., and Lawrence W. Kenny. "Individual teacher incentives and student performance." *Journal of Public Economics* 91, no. 5-6 (2007): 901-914.

Geiser, Saul, and Veronica Santelices. *The Role of Advanced Placement and Honors Courses in College Admissions*. Reserch and Occasional Paper Series, CSHE.4.04, Center for Studies on Higher Education, 2004.

Glewwe, P., N. Ilias, and M. Kremer. "Teacher Incentives." *National Bureau of Economic Research Working Paper Number 9671*, 2003.

Hastings, Justine S., Thomas Kane, and Douglas Staiger. "Gender, Performance and Preferences: Do Girls and Boys Respond Differently to School Environment? Evidence from School Assignment by Randomized Lottery." *American Economic Review Papers and Proceedings* 96, no. 2 (2006): 232-236.

Hauser, Robert. *Trends in College Entry among Whites, Blacks, and Hispanics*. Edited by Charles T. Clotfelter and Michael Rothschild. Vol. Studies of Supply and Demand in Higher Education. Chicago: University of Chicago Press, 1993.

Holmstrom, Bengt, and Paul Milgrom. "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design." *Journal of Law, Economics and Organization* 7, no. Special Issue (1991): 24-52.

Horn, L. *Stopouts or stayouts? Undergraduates who leave college in their first year*. (NCES 1999-087), Washington, DC: U.S. Department of Education, NCES. U.S. Government Printing Office., 1998.

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

Jackson, C. Kirabo. "A little now for a lot later: A look at a Texas advanced placement incentive program." *Journal of Human Resources* 45, no. 3 (2010).

Jackson, C. Kirabo. "Ability-grouping and Academic Inequality: Evidence from Rule-Based Student Assignments." *National Bureau of Economic Research Working Paper*, 2009.

Jackson, C.K. (2010). 'Do Students Benefit from Attending Better Schools?: Evidence from rule-based student assignments in Trinidad and Tobago', *The Economic Journal*, doi: 10.1111/j.1468-0297.2010.02371.x.

Jackson, C. Kirabo, and Elias Bruegmann. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics* 1, no. 4 (2009): 85-108.

Jacob, Brian A., and Steven D Levitt. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118, no. 3 (2003): 843-877.

Jencks, C., and M. Phillips. *The black-white test score gap*. Washington, DC: Brookings Institution Press, 1998.

Kalsner, L. "Issues in College Student Retention." *Higher Education Extension Service Review*, 1991.

Kane, Thomas. "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education." *Journal of Political Economy* 102 (October 1994): 878-911.

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

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F Katz. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75, no. 1 (2007): 83-119.

Klopfenstein, K. "The advanced placement expansion of the 1990s: How did traditionally underserved students fare?" *Education Policy Analysis Archives* 12, no. 68 (2004).

Kohn, A. *Punished by rewards the trouble with gold stars, incentive plans, A's, praise, and other bribes*. Bridgewater, NJ: Replica Books., 1999.

Kremer, Michael, Edward Miguel, Rebecca Thornton, and Owen Ozier. "Incentives to Learn." *World Bank Policy Research Working Paper 3546*, 2004.

Lavy, Victor. "Performance Pay and Teachers' Effort, Productivity and Grading Ethics." *American Economic Review*, December 2009.

Lee, David. S. "Trimming for Bounds on Treatment Effects with Missing Outcomes." *NBER Technical Working Paper T0277*, June 2002.

Lyon, J. "Program to offer students money for high test scores." *Arkansas News Bureau*, October 9, 2007.

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

Medina, J. "Making Cash a Prize for High Scores on Advanced Placement Tests." *The New York Times*, October 15, 2007.

Neal, Derek, and William Johnson. "The Role of Premarket Factors in Black-White Wage Differences." *The Journal of Political Economy* 104, no. 5 (1996): 869-895.

Scott-Clayton, Judith. "On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement." *unpublished mimeo*, 2008.

Seftor, Neil S., Arif Mamun, and Allen Schirm. "The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation." *Mathematica Report*, 2009.

Stinebrickner, Todd R., and Ralph Stinebrickner. "Learning about Academic Ability and the College Drop-out Decision." *NBER Working Paper 14810*, 2009.

Summers, Clyde W. "Preferential Admissions: An Unreal Solution to a Real Problem." *University of Toledo Law Review* 2, no. 2-3 (1970): 377-402.

Tienda, Marta, and Sunny Xinchun Niu. "Flagships, Feeders, and the Texas Top 10% Law: A Test of the "Brain Drain" Hypothesis." *The Journal of Higher Education* 77, no. 4 (2006).

Tinto, V. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Second Edition. Chicago: University of Chicago Press, 1993.

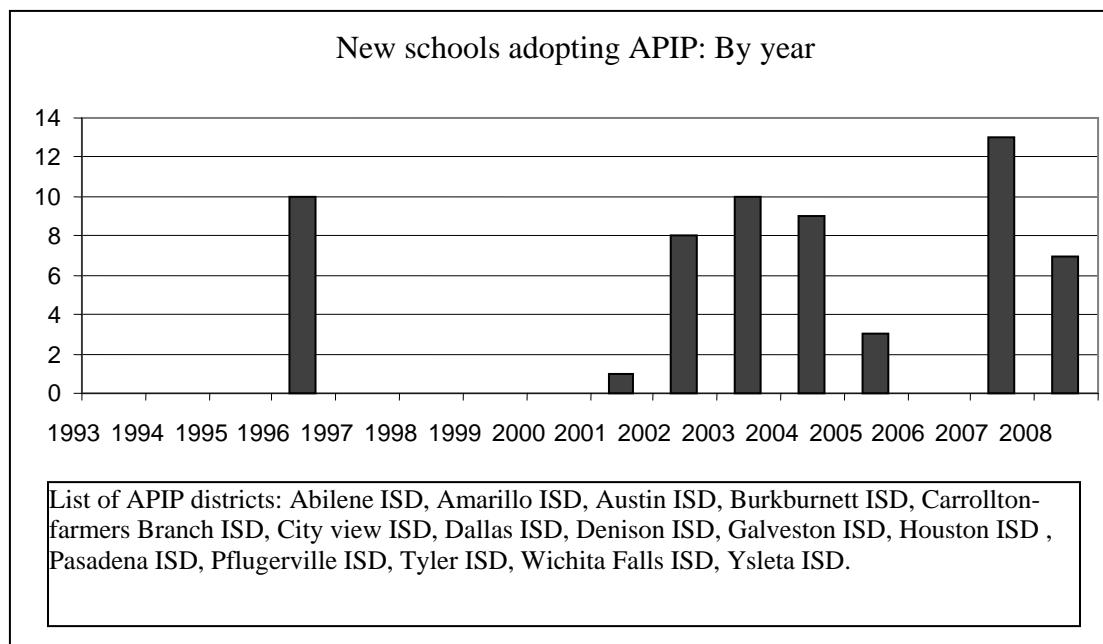
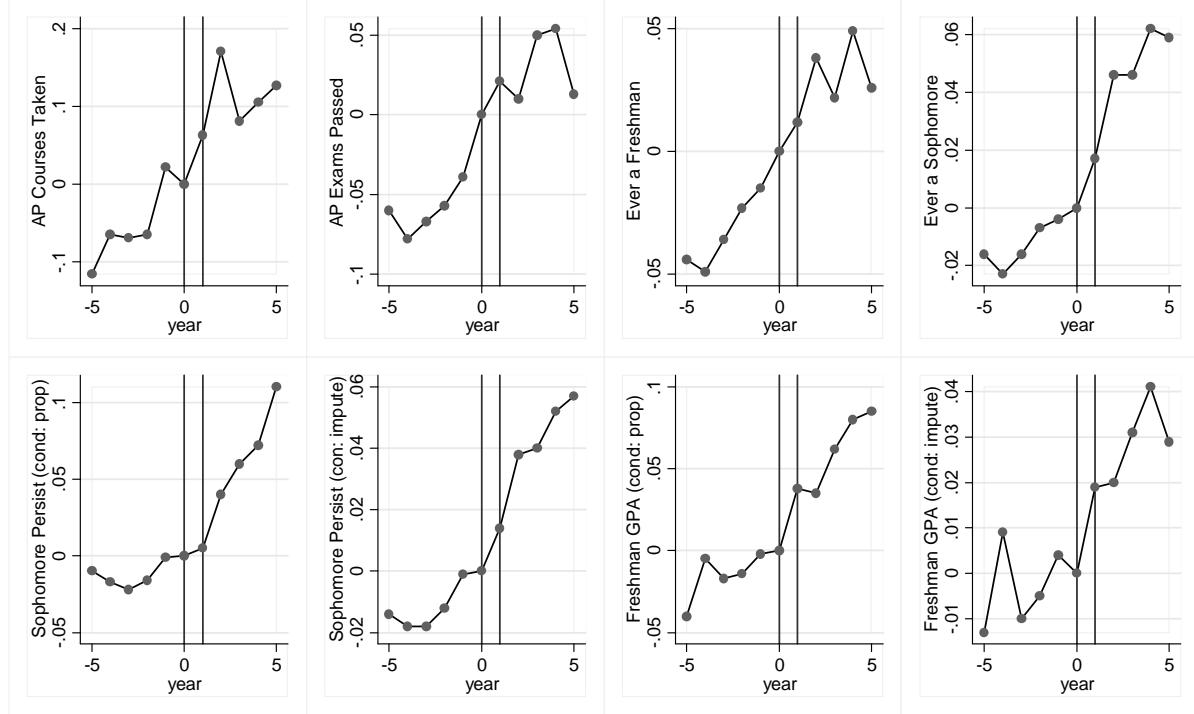


Figure 1: New APIP schools by year.

Dynamic Effects of APIP adoption: First affected cohort is year 0



Note: For all outcomes, the F-statistic associated with the null-hypothesis that the pre-treatment years differ from year t-1 yield p -values greater than 0.1 for all outcomes. In contrast the F-statistics associated with the null-hypothesis that the post treatment years differ from year t-1 yield p -values smaller than 0.05 for all outcomes.

Figure 2: Dynamic Effects of APIP Adoption.

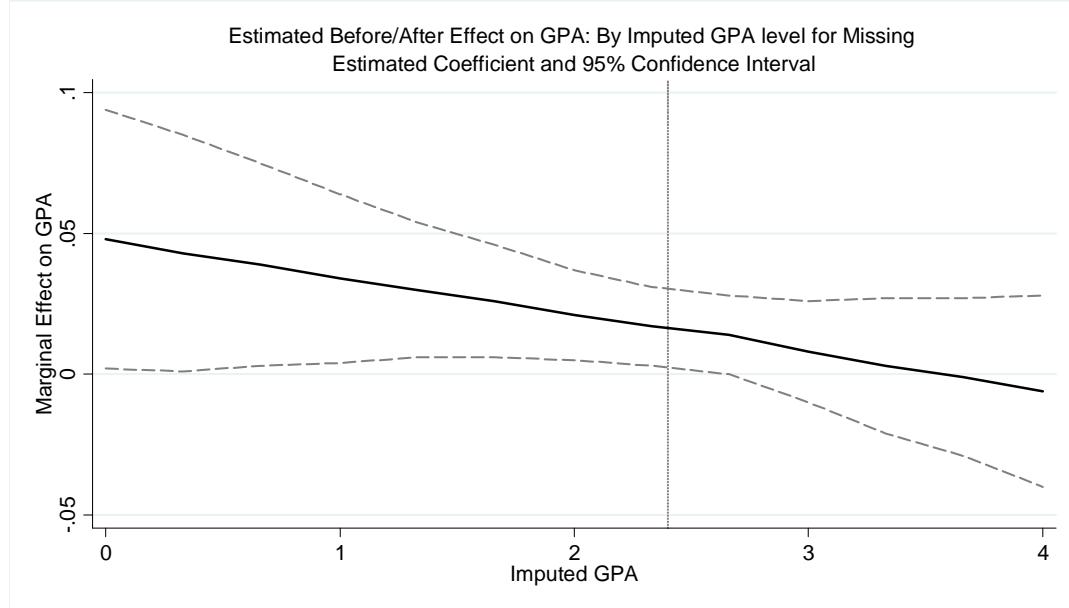


Figure 3: Sensitivity to Imputed GPA.

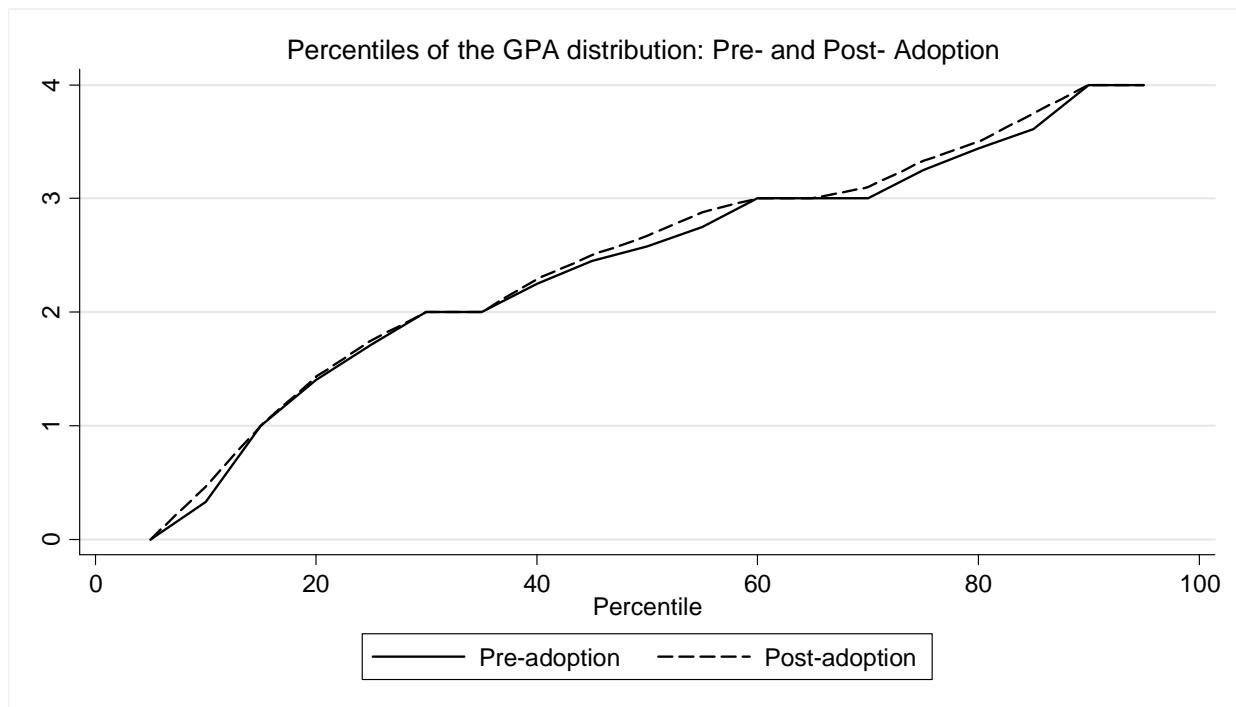


Figure 4: Distribution of GPA in pre- and post-adoption cohorts.

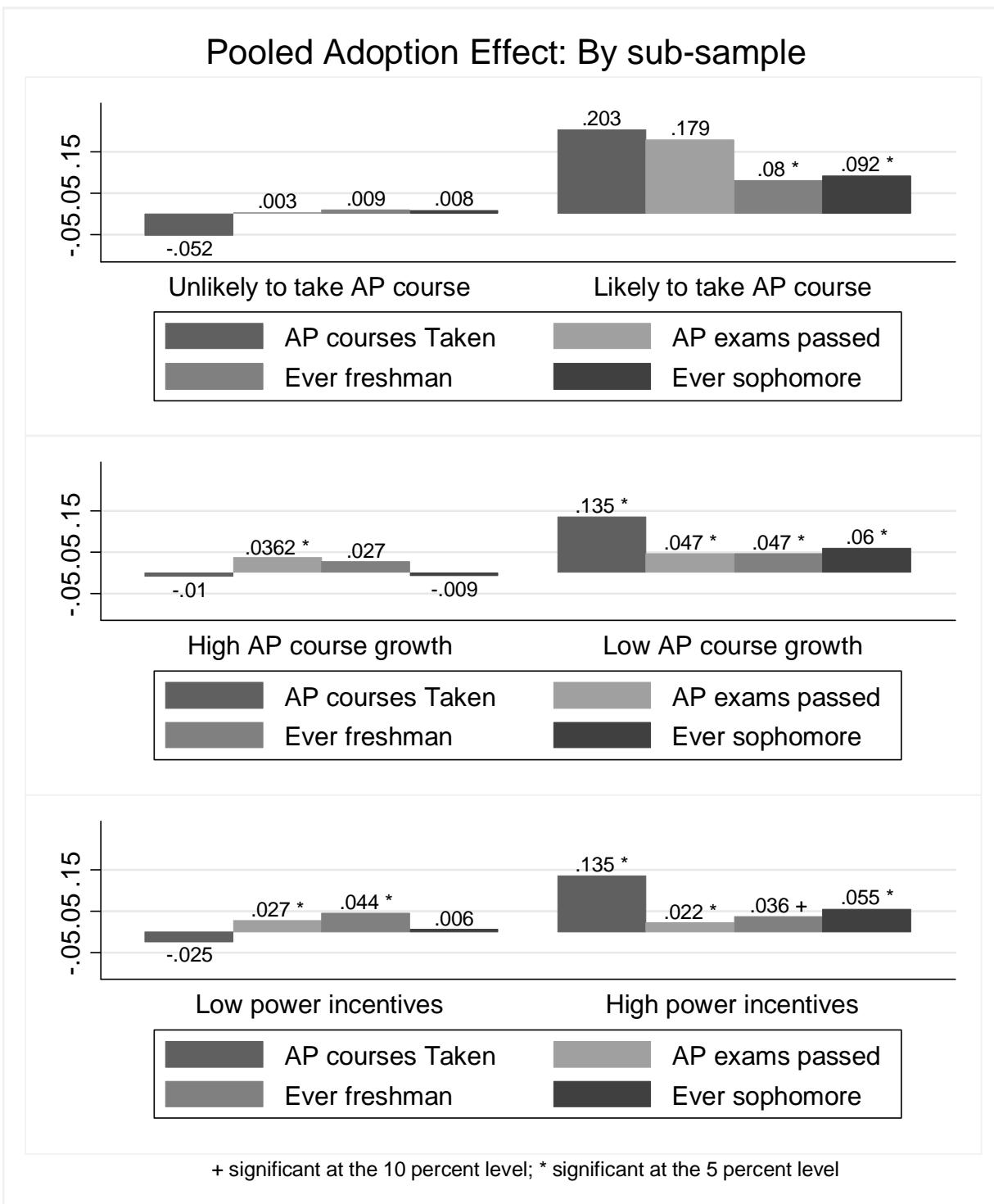


Figure 5: Effect of the APIP on different sub-samples

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 (642.34)	1836.36 (648.86)	716.85 (781.97)	751.56 (833.36)
% White	30.82 (25.43)	25.16 (23.28)	59.38 (29.46)	53.36 (30.42)
% Black	30.17 (26.82)	26.24 (23.5)	10.32 (15.64)	11.30 (17.08)
% Hispanic	35.76 (23.49)	45.36 (23.84)	28.92 (28.9)	33.67 (29.5)
% Asian	2.93 (3.43)	2.39 (3.65)	1.09 (2.76)	1.12 (2.98)
% Free lunch	34.33 (22.3)	41.60 (25.0)	30.42 (23.97)	35.51 (26.25)
% Limited English	9.66 (12.89)	10.68 (11.86)	3.57 (7.71)	3.83 (6.8)
City	0.874 (0.28)	0.739 (0.44)	0.182 (0.39)	0.197 (0.4)
Rural	0.000 (0.0)	0.017 (0.13)	0.489 (0.5)	0.373 (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)
10th Grade Reading z-Score	156858	-0.092	(1.018)	137704	-0.063	(0.987)
10th 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)
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)
Freshman year GPA	58685	2.382	(1.176)	44425	2.427	(1.192)
Graduate with a BA	147779	0.146	(0.354)	80931	0.115	(0.319)
Graduate with a AA	156858	0.038	(0.191)	115783	0.021	(0.142)
Graduate with a BA within 5	16871	0.019	(0.137)	71513	0.011	(0.104)
Graduate with a AA within 4	9079	0.004	(0.063)	56773	0.002	(0.042)
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)
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 2yr	156858	0.418	(0.493)	137704	0.392	(0.488)
Freshman within 0 years	156858	0.351	(0.536)	137704	0.411	(0.585)
Freshman within 1 year	156858	0.462	(0.627)	115783	0.504	(0.653)
Freshman within 3 years	147779	0.527	(0.669)	80931	0.561	(0.687)

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

Robust standard errors in brackets adjusted for clustering at the school level.

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

All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 4

Regression Estimates: Effect of years of APIP adoption on Freshman GPA and Enrolling in college as a sophomore

	1	2	3	4	5	6	7	8
	Freshman Year GPA			Ever a Sophomore				
	None	Imputation	Propensity score	Unconditional	Imputation	None	Propensity score	Trim: lower bound
Adopted (ITT year>0)	0.03 [0.018]	0.015 [0.007]*	0.03 [0.017]+	0.03 [0.009]**	0.026 [0.008]**	0.024 [0.016]	0.024 [0.016]	0.01 [0.017]
ITT year=1	0.009 [0.020]	0.004 [0.007]	0.009 [0.020]	0.008 [0.009]	0.007 [0.007]	0 [0.015]	-0.001 [0.015]	0.001 [0.016]
ITT year=2	0.043 [0.026]+	0.022 [0.010]*	0.044 [0.026]+	0.027 [0.010]**	0.023 [0.009]*	0.008 [0.019]	0.008 [0.019]	-0.004 [0.020]
ITT year=3	0.034 [0.032]	0.02 [0.014]	0.035 [0.032]	0.056 [0.012]**	0.047 [0.011]**	0.045 [0.021]*	0.045 [0.021]*	0.03 [0.022]
ITT year=4+	0.068 [0.029]*	0.031 [0.011]**	0.067 [0.029]*	0.06 [0.015]**	0.055 [0.012]**	0.086 [0.023]**	0.086 [0.023]**	0.066 [0.024]**
Cond. on Freshman School FX	YES	NO	YES	NO	NO	YES	YES	YES
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
F-test	0.068	0.035	0.073	>0.001	>0.001	>0.001	>0.001	>0.001
Observations	102,413	290,343	102,413	268,652	268,652	96,706	96,706	89,240

Robust standard errors in brackets adjusted for clustering at the school level.

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

All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 5

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 [0.015]	0.026 [0.017]	0.027 [0.018]	0.032 [0.018]+	0.034 [0.018]+	0.011 [0.008]	0.011 [0.009]	0.016 [0.009]+	0.017 [0.009]+
ITT year=2	0.036 [0.016]*	0.045 [0.017]*	0.05 [0.018]**	0.053 [0.018]**	0.054 [0.018]**	0.02 [0.010]+	0.021 [0.011]+	0.023 [0.011]*	0.024 [0.011]*
ITT year=3	0.038 [0.019]+	0.048 [0.021]*	0.055 [0.021]*	0.06 [0.021]**	0.063 [0.021]**	0.036 [0.010]**	0.042 [0.012]**	0.049 [0.012]**	0.052 [0.012]**
ITT year=4+	0.017 [0.024]	0.025 [0.024]	0.028 [0.024]	0.033 [0.024]	0.039 [0.023]+	0.051 [0.013]**	0.053 [0.016]**	0.057 [0.016]**	0.058 [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 adjusted for clustering at the school level.

+ significant at 10%; * significant at 5%; ** significant at 1%

All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 6

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
	Math Score	Read Score										
	10th Grade	10th Grade	LEP	Low Income	Black	Hispanic	Predicted: AP Courses	Predicted: AP Exams	Predicted: Attend College	Predicted: GPA	Predicted: GPA (enrollees)	Predicted: Sophomore
Adopted	-0.039	-0.013	0.009	-0.039	0.005	-0.025	0.01	-0.004	-0.001	-0.008	-0.013	0.002
ITT year>0	[0.025]	[0.021]	[0.011]	[0.018]*	[0.011]	[0.012]*	[0.032]	[0.040]	[0.006]	[0.010]	[0.012]	[0.007]
ITT year=1	-0.02	0.013	0.009	-0.021	0.003	-0.014	0.002	-0.007	-0.003	-0.011	-0.019	0
	[0.026]	[0.021]	[0.007]	[0.015]	[0.008]	[0.010]	[0.026]	[0.024]	[0.005]	[0.011]	[0.013]	[0.006]
ITT year=2	-0.051	-0.035	0.005	-0.044	0.005	-0.024	0.015	0.004	-0.005	-0.013	-0.013	0.005
	[0.032]+	[0.024]	[0.012]	[0.021]*	[0.011]	[0.013]+	[0.038]	[0.045]	[0.006]	[0.012]	[0.013]	[0.008]
ITT year=3	-0.049	-0.016	0.011	-0.049	0.001	-0.040	0.014	-0.024	-0.005	-0.001	-0.005	0.005
	[0.034]	[0.030]	[0.016]	[0.024]*	[0.015]	[0.017]*	[0.041]	[0.039]	[0.006]	[0.013]	[0.013]	[0.008]
ITT year=4+	-0.050	-0.047	0.017	-0.068	0.012	-0.045	0.018	0.017	-0.007	0.009	0.002	-0.002
	[0.033]	[0.29]	[0.022]	[0.030]*	[0.016]	[0.020]*	[0.048]	[0.052]	[0.007]	[0.014]	[0.017]	[0.010]
School FX	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
Observations	290343	290343	290343	290343	290343	290343	290343	290343	290343	290343	104367	290343
F: no effect	0.29	0.07	0.45	0.22	0.14	0.15	0.98	0.88	0.89	0.46	0.44	0.63

Robust standard errors in brackets adjusted for clustering at the school level.

+ significant at 10%; * significant at 5%; ** significant at 1%

Table 7

Summary Means and Standard Deviations for outcomes by Ethnicity

	Black		Hispanic		White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Take any AP course	0.225	(0.418)	0.211	(0.408)	0.354	(0.478)
AP courses Taken	0.638	(1.521)	0.560	(1.415)	1.186	(2.107)
Take any AP exams	0.034	(0.182)	0.039	(0.193)	0.106	(0.308)
AP Exams Taken	0.052	(0.333)	0.064	(0.393)	0.203	(0.756)
AP Exams Passed	0.009	(0.133)	0.021	(0.192)	0.113	(0.542)
Read score 10th grade	-0.257	(1.044)	-0.203	(0.995)	0.247	(0.903)
Math score 10th grade	-0.343	(0.966)	-0.203	(0.951)	0.261	(0.939)
LEP	0.008	(0.091)	0.258	(0.438)	0.006	(0.08)
Low income	0.476	(0.499)	0.574	(0.494)	0.146	(0.353)
Female	0.521	(0.5)	0.505	(0.5)	0.499	(0.5)
Ever a freshman	0.566	(0.689)	0.445	(0.618)	0.773	(0.722)
Ever a sophomore	0.256	(0.523)	0.230	(0.494)	0.445	(0.625)
Ever a junior	0.119	(0.327)	0.095	(0.296)	0.266	(0.448)
Freshman GPA	2.071	(1.204)	2.319	(1.187)	2.599	(1.131)
Graduate with BA	0.090	(0.286)	0.080	(0.271)	0.232	(0.422)
Graduate with AA	0.019	(0.137)	0.029	(0.167)	0.041	(0.198)
College outside TX	0.050	(0.217)	0.018	(0.134)	0.071	(0.256)
College in TX	0.383	(0.486)	0.383	(0.486)	0.526	(0.499)
Freshman at Private college	0.047	(0.212)	0.013	(0.114)	0.067	(0.251)
Freshman at 4yr College	0.187	(0.4)	0.099	(0.302)	0.269	(0.454)
Observations	69445		128291		83505	

Table 8

Regression Estimates: APIP Effects by cohort on College outcomes [interacted with demographic indicators]

	1	2	3	4	5	6	7	8		
	AP Courses Taken	AP Exams Taken	AP Exams Passed	Freshman ever	Freshman at 4yr	GPA	GPA	Sophomore ever	Sophomore ever	Sophomore ever
ITT year=1	-0.0378 [0.0910]	0.047 [0.033]	0.03 [0.025]	0.026 [0.0201]	0.0171 [0.0140]	0.00645 [0.0163]	0.029 [0.0315]	0.0233+ [0.0123]	0.0333** [0.0104]	0.0380* [0.0186]
* Black	0.039 [0.0893]	0.001 [0.039]	-0.013 [0.027]	0.00849 [0.0264]	0.0111 [0.0149]	0.00954 [0.0238]	0.00606 [0.0473]	-0.00465 [0.0137]	-0.0125 [0.0110]	-0.00866 [0.0223]
* Hispanic	-0.00514 [0.0800]	0.025 [0.031]	0.005 [0.025]	-0.0274 [0.0194]	-0.0166 [0.0124]	-0.0241 [0.0202]	-0.0614 [0.0498]	-0.0361* [0.0142]	-0.0471** [0.0116]	-0.0742** [0.0193]
* Female	-0.00795 [0.0229]	0.005 [0.010]	0.001 [0.008]	0.0230* [0.00951]	0.00923 [0.00582]	0.0115 [0.0118]	0.00789 [0.0304]	0.00699 [0.00919]	-0.00143 [0.00812]	-0.00584 [0.0149]
ITT year=2	-0.0534 [0.0910]	0.087 [0.020 **]	0.062 [0.014 **]	0.0450* [0.0218]	0.024 [0.0167]	0.00E+00 [0.0194]	0.0467 [0.0407]	0.0260+ [0.0131]	0.0252* [0.0118]	-0.0138 [0.0214]
* Black	0.0238 [0.0804]	-0.002 [0.019]	-0.029 [0.013]*	0.00906 [0.0269]	0.0179 [0.0196]	0.0232 [0.0264]	-0.00487 [0.0661]	-0.00126 [0.0152]	-0.000174 [0.0129]	0.0234 [0.0243]
* Hispanic	0.0835 [0.0787]	0.021 [0.021]	-0.012 [0.015]	-0.0294 [0.0199]	-0.0133 [0.0148]	0.0281 [0.0204]	-0.00271 [0.0771]	-0.00782 [0.0147]	-0.0152 [0.0141]	0.0205 [0.0314]
* Female	0.0828** [0.0290]	-0.026 [0.011]*	-0.017 [0.007]*	0.0124 [0.00899]	0.000786 [0.00540]	0.015 [0.0135]	0.0154 [0.0326]	0.0156 [0.00988]	0.0137 [0.00936]	0.0422+ [0.0223]
ITT year=3	0.128 [0.0916]	0.086 [0.029 **]	0.046 [0.016 **]	0.0781** [0.0262]	0.0376+ [0.0205]	0.0144 [0.0216]	0.0257 [0.0427]	0.0692** [0.0209]	0.0657** [0.0173]	0.0458+ [0.0258]
* Black	-0.0712 [0.0719]	-0.039 [0.031]	-0.032 [0.018]+	0.00936 [0.0271]	0.0206 [0.0214]	0.0499+ [0.0265]	0.0749 [0.0569]	-0.00968 [0.0221]	-0.00834 [0.0188]	0.000164 [0.0300]
* Hispanic	0.0487 [0.0744]	-0.028 [0.031]	-0.016 [0.018]	-0.0652** [0.0211]	-0.0193 [0.0169]	0.00855 [0.0228]	0.0446 [0.0649]	-0.0291 [0.0185]	-0.0370* [0.0169]	-0.00607 [0.0343]
* Female	0.0349 [0.0301]	0.011 [0.010]	0.007 [0.008]	0.0138 [0.0101]	-0.00282 [0.00688]	-0.0195 [0.0167]	-0.023 [0.0369]	0.00121 [0.00896]	-0.00345 [0.00863]	0.0173 [0.0170]
ITT year=4+	0.138 [0.128]	0.104 [0.025 **]	0.045 [0.014 **]	0.0534+ [0.0297]	0.0265 [0.0256]	-0.0102 [0.0156]	0.0358 [0.0366]	0.0531* [0.0210]	0.0482** [0.0159]	0.0485* [0.0231]
* Black	-0.0833 [0.0723]	-0.012 [0.019]	-0.002 [0.012]	0.00976 [0.0300]	0.0333 [0.0257]	0.138** [0.0156]	0.161** [0.0452]	0.0272 [0.0181]	0.0417** [0.0129]	0.0857** [0.0207]
* Hispanic	-0.0492 [0.0611]	-0.004 [0.019]	0.007 [0.012]	-0.0543* [0.0270]	-0.0115 [0.0219]	0.0118 [0.0147]	0.0311 [0.0474]	-0.00187 [0.0186]	-0.00678 [0.0142]	0.0494* [0.0218]
* Female	-0.0531* [0.0262]	0.006 [0.008]	0.002 [0.004]	0.0289** [0.00695]	0.000277 [0.00417]	0.00831 [0.00965]	0.00258 [0.0219]	0.00749 [0.00862]	0.000187 [0.00779]	0.00154 [0.0124]
Pr(all ethnic same)	0.14	0.75	0.06	>0.000	0.12	>0.000	0.018	0.002	>0.000	>0.000
Pr(all sex same)	0.002	0.139	0.154	0.0013	0.579	0.151	0.853	0.519	0.341	0.267
Year	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School-by-race FX	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Control for selection	N/A	N/A	N/A	N/A	N/A	impute	prop	uncond.	impute	prop
Observations	290,343	290,343	290,343	290,343	290,343	290,343	102,413	268,652	268,652	96,706

Robust standard errors in brackets adjusted for clustering at the school level.

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

All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 9

Regression Estimates: Effect of Years of APIP adoption on longer-run College outcomes by Ethnicity

	1	2	3	4	5	6	7	8	9
	Any degree	Any degree within 4 years	BA	Total Credits	Total Credits within 4 years	Ever a Junior	Junior within 4 years	Ever a sophomore	Ever a freshman
All (224971 observations)									
ITT year= 1	0.005 [0.006]	0.002 [0.003]	0.008 [0.006]	4.395 [1.843]*	1.589 [2.022]	0.016 [0.006]**	0.012 [0.006]+	0.027 [0.010]*	0.033 [0.017]+
ITT year= 2	0.003 [0.009]	0 [0.004]	0.008 [0.009]	6.626 [1.708]**	2.606 [2.000]	0.019 [0.007]**	0.013 [0.007]+	0.033 [0.010]**	0.053 [0.017]**
ITT year= 3	0.005 [0.012]	0 [0.004]	0.009 [0.012]	9.368 [2.205]**	5.061 [2.209]*	0.022 [0.009]*	0.015 [0.009]+	0.061 [0.013]**	0.065 [0.019]**
ITT year= 4+	-0.018 [0.016]	-0.01 [0.006]+	-0.016 [0.016]	4.185 [2.413]+	3.708 [2.457]	0.013 [0.011]	0.008 [0.010]	0.065 [0.016]**	0.04 [0.023]+
Black (54059 observations)									
ITT year= 1	0.016 [0.010]	0.002 [0.005]	0.018 [0.009]*	3.131 [2.819]	0.055 [2.190]	0.016 [0.009]+	0.013 [0.008]+	0.011 [0.014]	0.001 [0.026]
ITT year= 2	0.01 [0.011]	-0.003 [0.005]	0.011 [0.011]	3.461 [2.247]	-0.775 [1.595]	0.011 [0.008]	0.004 [0.006]	0.011 [0.013]	0.013 [0.022]
ITT year= 3	0.017 [0.012]	-0.005 [0.005]	0.018 [0.011]+	7.68 [2.600]**	3.525 [1.765]+	0.017 [0.010]+	0.012 [0.008]	0.034 [0.014]*	0.038 [0.021]+
ITT year= 4+	0.003 [0.012]	-0.01 [0.007]	0.010 [0.011]	5.997 [3.674]	2.94 [1.761]	0.012 [0.012]	0.01 [0.010]	0.047 [0.020]*	-0.012 [0.025]
Hispanic (93171)									
ITT year= 1	0.011 [0.006]+	0 [0.005]	0.012 [0.005]*	1.68 [2.025]	-1.998 [2.430]	0.013 [0.006]*	0.004 [0.006]	0.021 [0.013]	0.015 [0.016]
ITT year= 2	0.018 [0.010]+	-0.003 [0.006]	0.022 [0.008]*	4.561 [2.372]+	-1.187 [3.039]	0.022 [0.008]**	0.011 [0.007]	0.035 [0.014]*	0.03 [0.018]
ITT year= 3	0.024 [0.010]*	-0.005 [0.005]	0.026 [0.010]*	6.828 [1.949]**	-0.549 [2.349]	0.021 [0.007]**	0.007 [0.006]	0.052 [0.014]**	0.027 [0.021]
ITT year= 4+	0.023 [0.012]+	-0.003 [0.007]	0.025 [0.011]*	4.793 [2.878]	-1.579 [3.518]	0.021 [0.009]*	0.008 [0.009]	0.062 [0.019]**	0.016 [0.028]
White (68722)									
ITT year= 1	0.004 [0.009]	0.004 [0.007]	0.01 [0.009]	4.909 [3.004]	2.552 [2.659]	0.024 [0.011]*	0.023 [0.011]*	0.038 [0.015]*	0.041 [0.024]+
ITT year= 2	0.004 [0.011]	0.003 [0.006]	0.013 [0.012]	6.505 [2.679]*	3.417 [2.627]	0.024 [0.012]+	0.023 [0.013]+	0.042 [0.017]*	0.064 [0.025]*
ITT year= 3	-0.005 [0.012]	0.005 [0.007]	0.005 [0.013]	9.038 [2.763]**	5.275 [3.129]+	0.027 [0.016]+	0.027 [0.017]	0.075 [0.021]**	0.078 [0.029]**
ITT year= 4+	-0.017 [0.018]	-0.007 [0.010]	-0.011 [0.018]	7.107 [2.815]*	5.421 [2.804]+	0.024 [0.017]	0.026 [0.017]	0.089 [0.023]**	0.074 [0.034]*

Robust standard errors in brackets adjusted for clustering at the school level.

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

All regressions control for 10th grade test scores, gender, ethnicity, LEP status and free or reduced lunch status.

Appendix

Appendix Table 1: Robustness Checks

Regression Estimates: Effect of APIP adoption years on college outcomes for different sub-samples (robustness check)									
	1	2	3	4	6	7	8	9	10
	AP Courses Taken	AP Exams Taken	AP Exams Passed	Ever a freshman	GPA imputed	GPA (cond.)	Ever sophomore	Ever sophomore (imputed)	Ever sophomore (cond.)
No controls									
ITT year= 1	-0.017 [0.051]	0.065 [0.012]**	0.031 [0.008]**	0.029 [0.017]+	0.002 [0.012]	0.003 [0.022]	0.007 [0.011]	0.006 [0.010]	-0.003 [0.016]
ITT year= 2	0.034 [0.063]	0.883 [0.015]**	0.041 [0.008]**	0.034 [0.020]+	0.014 [0.014]	0.039 [0.028]	0.024 [0.012]*	0.019 [0.012]	0.004 [0.019]
ITT year= 3	0.123 [0.076]	0.064 [0.017]**	0.028 [0.009]**	0.061 [0.023]*	0.013 [0.016]	0.023 [0.032]	0.052 [0.014]**	0.044 [0.014]**	0.043 [0.021]*
ITT year= 4+	0.059 [0.108]	0.095 [0.020]*	0.043 [0.011]**	0.046 [0.025]+	0.034 [0.016]*	0.071 [0.031]*	0.059 [0.016]**	0.055 [0.014]**	0.087 [0.022]**
Middle school and high school fixed effect and only feeder schools									
ITT year= 1	0.096 [0.052]+	0.075 [0.018]**	0.039 [0.012]**	0.013 [0.015]	0.004 [0.009]	0.011 [0.021]	-0.009 [0.014]	-0.005 [0.011]	-0.015 [0.020]
ITT year= 2	0.166 [0.065]*	0.085 [0.021]**	0.046 [0.011]**	0.037 [0.017]*	0.03 [0.013]*	0.045 [0.035]	0.01 [0.016]	0.009 [0.013]	-0.007 [0.024]
ITT year= 3	0.3 [0.080]**	0.086 [0.023]**	0.041 [0.012]**	0.054 [0.017]**	0.026 [0.016]	0.031 [0.038]	0.047 [0.014]**	0.039 [0.012]**	0.035 [0.023]
ITT year= 4+	0.172 [0.083]*	0.122 [0.031]**	0.06 [0.017]**	0.052 [0.022]*	0.027 [0.012]*	0.049 [0.028]+	0.047 [0.018]*	0.04 [0.014]**	0.047 [0.027]+
School intercepts and trends included									
ITT year= 1	0.003 [0.050]	0.065 [0.016]**	0.034 [0.010]**	0.008 [0.011]	0.004 [0.009]	-0.008 [0.024]	-0.001 [0.010]	0.001 [0.008]	-0.007 [0.015]
ITT year= 2	0.074 [0.062]	0.084 [0.016]**	0.045 [0.009]**	0.025 [0.015]+	0.024 [0.014]	0.01 [0.048]	0.004 [0.012]	0.003 [0.010]	-0.021 [0.017]
ITT year= 3	0.201 [0.082]*	0.08 [0.022]**	0.04 [0.011]**	0.03 [0.017]+	0.049 [0.017]**	0.029 [0.055]	0.03 [0.014]*	0.026 [0.010]*	0.018 [0.024]
ITT year= 4+	0.113 [0.105]	0.129 [0.031]**	0.065 [0.017]**	0.017 [0.021]	0.078 [0.016]**	0.075 [0.060]	0.028 [0.016]+	0.026 [0.014]+	0.035 [0.034]

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

Appendix Table 2: 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 ITT year>0	0.004 [0.059]	0.058 [0.066]	0.003 [0.046]	0.026 [0.055]	0.004 [0.068]	-0.092 [0.070]
Observations	411	466	473	583	530	580
R-squared	0.12	0.1	0.13	0.1	0.11	0.12

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

* significant at 5% level; ** significant at 1% level. All regressions include school and year fixed effects.

Appendix Table 3: 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		Social Sciences and History		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% level; ** significant at 1% level. All regressions include school and year fixed effects.

Appendix Table 4: 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 10th Grade	School Rank in 10th Grade	Normalized 10th grade Math Score	Normalized 10th grade Reading Score	Predicted GPA	School Rank in 10th Grade	School Rank in 10th Grade	Normalized 10th grade Math Score	Normalized 10th 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 parenthesis are adjusted for clustering at the school level.

* significant at 5% level; ** significant at 1% level

Appendix Table 5

	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 [0.01]	0.014 [0.013]	-0.021 [0.077]	-0.029 [0.025]	-0.054 [0.027]+	-0.418 [0.455]	-0.839 [2.206]	0.307 [0.410]	-0.147 [1.004]
ITT years= 2	0.016 [0.011]	0.013 [0.014]	0.161 [0.107]	-0.061 [0.041]	-0.089 [0.053]	1.118 [0.633]	-7.308 [3.090]*	0.486 [0.545]	2.127 [0.897]*
ITT years= 3	0.015 [0.01]	0.012 [0.016]	-0.071 [0.073]	-0.088 [0.035]*	-0.046 [0.042]	1.747 [1.231]	-3.406 [2.526]	0.117 [0.598]	0.439 [1.384]
ITT years= 4+	0.014 [0.012]	0.018 [0.017]	-0.07 [0.067]	0.074 [0.037]	0.03 [0.052]	2.634 [1.017]*	5.624 [2.953]	-0.101 [0.537]	-1.002 [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% level; ** significant at 1% level.

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 analysis are based on within-school variation, this kind of self-selection will not bias the result as long as there are not 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 kinds of schools with relationships to donor organizations. However, is does not imply that the timing of adoption is endogenous to changes within schools. Of the schools that were treated in sample, 7 of them had donors for which there were previous projects.⁴⁴ 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 are unchanged. Of the comparison APIP schools [i.e. those APIP school that have not yet adopted], 7 of the Austin schools had relationships with the Michael and Susan Dell Foundation. These 7 Austin schools adopt the APIP in 2008, so they serve as comparison schools in my data. However, starting in 2002 there was the Advancement Via Individual Determination (AVID) program, Project Advance, and Project Smart in these schools. All of these programs are College readiness programs that would lead to an *underestimate* of the APIP effect [because these schools serve as comparison schools rather than 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 5 of the

⁴⁴ Dodge Jones Foundation in Abilene (2 in 2003); Perkins Prothro Foundation in Wichita Falls (3 in 2002); Munson Foundation in Denison (1 in 2004); Fourth Partner Foundation in Tyler (2 in 2002).

Dallas schools that started the program in 2003⁴⁵, 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 5 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 the 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 the 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 APIP was not adopted by the selected APIP schools. However all the estimated effects are above and beyond any effect from the statewide program. [Source: Texas Education Agency Press Release: “Number of Advanced Placement Exams Taken by Texas Students Increases Dramatically”. August 23, 2000].

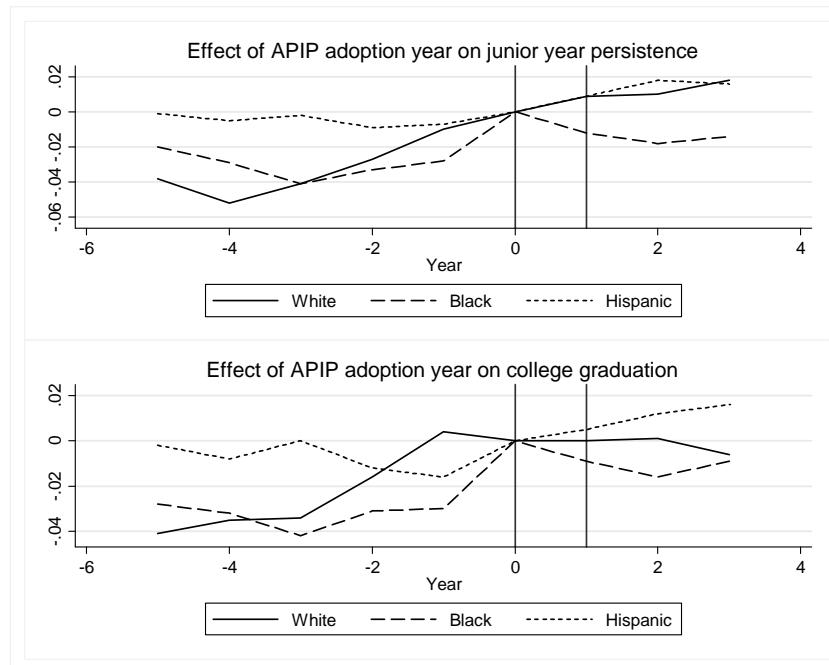


Figure A1: Effect of the APIP on persisting to Junior Year and Graduating by Ethnicity

⁴⁵ Kimball, Roosevelt, Sunset, Thomas Jefferson, Seagoville high Schools.