"A Stitch in Time: The Effects of a Novel Incentive-Based High-School Intervention on College Outcomes"[Ψ]

C. Kirabo Jackson   *October 16, 2009*

Abstract: I analyze the longer-run effects of a program that pays both 11th and 12th grade students and teachers for passing scores on Advanced Placement exams. Using a difference-in-differences strategy, I find that affected students 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 ─ groups that tend to underperform in college. This evidence suggests that relatively late high-school interventions may confer lasting positive and large effects on student achievement in college, 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.

Across the United States, college matriculation and completion rates for low-income and ethnic-minority students are much lower than those for non-poor whites.[1] These disparities are sobering given findings that 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). (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 ─ suggesting that policies that improve scholastic ability among low-income and under-represented ethnic minority students may reduce these differences in college-going.[2] While there are large differences in college going across groups, much of the gaps in educational attainment across ethnic and socioeconomic groups occur among those who enter college but do not persist (Adelman 1999, Bowen and Bok 1998, Jencks and Phillips 1998).[3] It is well documented that the majority of college attrition occurs in

Ψ 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, Larry Katz, and participants at the Dartmouth College Economics Seminar. All errors are my own.

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 implied two or four-year college completion rates for these same groups are 68% for whites, 51% for blacks and 53% for Hispanics.

2 Earlier research by (Hauser 1993, Kane 1994) find that differences in college-going rates across ethnic groups are related to differences in the ability to pay for college. However this in no longer accepted wisdom.

3 Among students enrolled at 4-year colleges in 2001, 59.4 percent of whites graduated with a degree within six years compared to only 46.8 for Hispanic students and 40.5 for black students [NCES, IPEDS Graduation Rate Survey]

the first year - so that persistence through the first year is a key predictor of subsequent college success (Brawer 1996, Horn 1998, Bradburn 2002). Since sufficient academic preparedness is key to the successful integration into college life (Tinto 1993, Kalsner 1991) policies that improve scholastic ability *before college entry* may increase the college persistence and graduation rates of students who would have enrolled in college even in the absence of such policies in addition to increasing college enrollment of marginal college enrollees.

Early educational interventions have been found to have large effects on adult outcomes (Currie 2001, Deming 2009) and it is argued that that remediation of inadequate early investments is difficult and costly (Cunha and Heckman 2007, Cunha, Heckman and Lochner 2006). However, I contend that if academic underperformance is a result of some economic inefficiency (such as imperfect information, student myopia, suboptimal teacher effort or suboptimal student effort), late interventions that alleviate such inefficiency could be very cost effective. Despite this possibility, there is little empirical evidence on the efficacy of late interventions. While there are numerous programs aimed at high school students with the aim of increasing college going and improving college readiness, rigorous evaluations of these interventions on college outcomes are lacking.[4] Because there is often no exogenous variation in pre-college student characteristics (because students may self-select into college preparation programs such as Advanced Placement and the International Baccalaureate programs) most studies on the efficacy of pre-college interventions on college success are largely descriptive.[5]

I aim to provide some of the first credible evidence on the efficacy of late high-school interventions. Specifically, I aim to determine the effects of a novel high-school intervention that includes cash incentives for both high-school teachers and high-school students for each passing score earned on an AP exam on students' college outcomes in addition to teacher training, curricular oversight, and test-prep sessions. The Dallas-based Advanced Placement Incentive Program (APIP) is targeted primarily to low-income, minority-majority school districts with a view towards improving college readiness. Due to the perceived success of this program, New Mexico and New York City have adopted similar programs while schools in Arkansas, Alabama, Connecticut, Kentucky, Massachusetts, and Virginia have replicated the APIP ─ so that the APIP

[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] In one notable exception, 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.

may be worthy of study in and of itself.[6] Using school-level data, (Jackson 2010) finds that the APIP increases AP participation, improves SAT and ACT performance and increases college enrollment. Since the stated aims of the APIP are to improve college readiness, it is important to determine not just whether the program improves high school outcomes, but whether the APIP improves performance for those students who enroll in college.

To this aim, I investigate how the APIP, which affected 11th and 12 grade students, affected (1) their likelihood of attending college (2) their likelihood of persisting through their sophomore and junior years (3) their college GPA and (4) the likelihood of graduating with a degree. A key feature of my analysis is that I look at the effect on college outcomes both overall *and for those students who would have enrolled in college absent the APIP*. I link all Texas students' high-school data to administrative college records ─ allowing me compare the college outcomes of students exposed to the APIP to those of students not exposed to the APIP as long as they attended any college in Texas. 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). Since the administrators of the APIP could not roll out the program to all interested high-schools at once due to a relative shortage of donors, there is variation in the timing of APIP adoption within the sample of interested schools. This allows me to use a difference-in-difference strategy ─ comparing the difference in outcomes between students with the same pre-treatment test scores from the same high-school before and after APIP adoption (i.e. exposed cohorts to not exposed cohorts) to the difference in outcomes between students with the same incoming test scores from other high-schools over the same time period. By comparing cohorts from the same high-school, I remove (1) selection *within* a cohort (the differences in unobserved attributes that make one student take AP courses while another does not) and (2) selection *across* schools (the differences in unobserved attributes that make students from certain schools more likely to excel at college than others). By using changes in outcomes for similar schools over the same time period as a comparison, I remove the effects of policies and factors that affect all schools and may coincide with APIP adoption at some schools.

While the difference-in-differences strategy removes several sources of confounding variation, there are three remaining endogeneity concerns; (1) The first concern is that APIP adoption may be endogenous. To deal with the first concern, I limit the estimation sample to only

[6] (Lyon 2007, Medina 2007, Mathews 2004), http://www.nationalmathandscience.org

those schools that ever adopt the APIP, with similar levels of motivation and interest in the APIP. The timing of APIP adoption, within this subsample, is determined by the idiosyncratic preferences and the exogenous availability of private donors. Supporting this assertion, I show that placebo treatments from before APIP adoption have no predictive power on the outcomes, and I show that the results are robust to including school specific linear trends. (2) The second concern is that students may self-select into treatment schools. I address this issue in three ways. First, instead of using the actual treatment, I define "intention to treat" based on a student's school enrollment in 10th grade rather than their school in 11th and 12th grade. Second, I show that the results are robust to eliminating all students who did not attend a feeder middle-school to the treatment high-schools, and making inferences within groups of students who attended the same middle-school and the same high-school. Lastly, I test for selective migration directly and show that students who were in the treatment cohorts actually had slightly *worse* incoming 10th grade test scores than before adoption – suggesting that any bias due to selective migration would be *negative*. (3) The last concern is that with data on Texas colleges only, *increased* shifting of college going from out-of-state to in-state as a result of the APIP would look like improvements in outcomes. To address this concern, I show that the incidence of out-of-state college going among the APIP schools is too small to generate the effects estimated, show that the APIP improves outcomes conditional on college enrolment, and show that this scenario is inconsistent with the data ─ so I am able to rule out the possibility that this affects the findings in any meaningful way. In sum, I am reasonably confident that any positive effects associated with APIP adoption reflect a real causal effect.

The APIP combines additional resources with financial incentives to induce students and teachers to use them. While incentives schemes in secondary school may improve students' contemporaneous outcomes and may increase students' likelihoods of attending university or college, it is unclear whether these schemes cause students to perform better *after* they enroll in college.[7] First, some psychologists argue that external rewards for children can supplant intrinsic motivation, such that effort and performance may be worse after the incentives are removed than

[7] (Angrist and Lavy forthcoming) 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. Looking at college interventions, (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 are more effective than performance incentives given to parents when the parents are less productive. In a related literature, (Dynarski 2008) (Scott-Clayton 2008) study the incentive effects of grade-contingent scholarships among college students.

if they had never been introduced (Kohn 1999). Second, improvements in outcomes may reflect test-taking effort so that the contemporaneous gains in test scores may not persist.[8] Third, marginal college enrollees may subsequently fail or drop-out if they are not sufficiently college ready. Fourth, a principal agent multitask model (Holmstrom and Milgrom 1991) predicts that improved AP performance could come at the expense of other important unrewarded skills. Given these reasons, and concerns that rewarding test performance may lead to undesirable practices such as "teaching to the test" and cheating, it is important to study the longer-term effects of incentive-based-interventions. This study might shed light on these issues.

Consistent with (Jackson 2010), affected students were more likely to take and pass AP exams, and more likely to enroll in college. Guidance counselors credit the 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. Furthermore, *conditional on college enrollment*, affected students had higher grades and increased persistence. While there were small gender differences, improvements were particularly pronounced for black and Hispanic students. Consistent with finding larger GPA and persistence effects for minority students, suggestive results on college graduation show that treated black and Hispanic students were more likely to graduate with a four-year degree ─ there was no such increase for white students.

This is one of the first papers to present compelling evidence that a high-school intervention may have lasting positive effects on student achievement, and may be effective at improving the outcomes of minority students. These findings also indicate that increasing participation in rigorous high-school programs such as APs can improve college readiness. Even though the APIP is not a pure cash incentive program, the results suggest that incentive programs may have lasting positive effects even after rewards are no longer provided. Lastly, these findings contribute to the debate on early vs. late interventions as they show that an inexpensive program targeted to high-school students can be very effective at increasing their eventual

[8] For example, 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 following the competition, 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, so that lack of persistence may not generalize to all contexts.

educational attainment.

The remainder of this paper is structured as follows. Section II presents a description of 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 checks, and robustness tests, shows that the results cannot be driven by students being more likely to enroll in Texas Colleges as a result of the APIP, and presents anecdotal evidence on possible mechanisms. Section VII concludes.

II. Description of the AP incentive program

AP courses are typically taken by students in 11th or 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.[9] The fact that selective colleges pay considerable 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, depending on the pace chosen by the teacher and the scope of the subject. 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 the type of cheating documented in (Jacob and Levitt 2003) unlikely. The exams are graded from 1 through 5, with 5 being the highest and 3 generally regarded as a passing grade. 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 for 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, an AP course substitutes for some activity that is less demanding.[10]

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

---

[9] While this is true in general, some highly selective colleges only allow students to use AP credits to pass out of prerequisites, but not towards regular graduation credit.
[10] **Source:** Executive Vice President AP Strategies and counselors at several Dallas high-schools.

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.[11] 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, 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 school, there is variation across schools in the amounts paid. Overall, the APIP can deliver a considerable increase in compensation for teachers.[12]

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.

[11] (Jackson and Bruegmann 2009) find that teacher learn from their peers so that vertical teams may be effective.

[12] 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).

As a general rule, adoption of the APIP works as follows. First, schools interested in implementing the APIP approach AP Strategies and are put on a list.[13] AP Strategies then tries to match interested schools to a donor. When a private donor approaches AP Strategies, he or she selects which schools to fund from within the group of willing schools. In most cases the donor wants a specific district.[14] 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.[15] It takes about two years to fully implement the APIP after a school expresses interest.

The donors choose the subjects that will be 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 55 schools that adopted the APIP between 1995 and 2007 (41 of which were early enough to have college outcomes) and 61 schools that had adopted the program by 2008. Since donors chose schools from within the group of willing schools, donor availability and donor preferences are the primary reasons for variation in the timing of program implementation. 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." Since several districts compete for the same donor, donor preferences determine the districts, or the schools within the district, that will adopt the program in any given year.[16] I argue that the exact timing of program adoption, within the group of willing schools, is orthogonal to *changes* in unobserved school characteristics. I test this assumption empirically in section VII.

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

[13] 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.

[14] For example: The first ten Dallas schools were chosen based on proximity to AP strategies; ST Microelectronics is located in the Carrolton-Farmers community and funded this district's schools; The Priddy Foundation specifically requested the Burkburnett and City View schools; anonymous donors specifically requested Amarillo and Pflugerville schools; The Dell foundation (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.

[15] 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.

[16] 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.

cost per student in an AP class ranges from $100 to $300. Private donors pay for between 60 to 75 percent of the total costs of the program, and the district covers the remainder. Districts typically pay for teacher training and corresponding travel, release time and some of the supplies and equipment costs. The donors fund the cash rewards to students and teachers, stipends to teachers for attending team meetings, bonuses to teachers and administrators for passing AP scores, and some of the supplies and equipment costs. Today, districts may be able to 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.

III. Theoretical Framework

In this section I provide a theoretical framework within which to interpret the empirical findings. Specifically, I discuss some of the theory on cash incentives in education, human capital theory, and the mismatch hypothesis and then discuss their implications for how the APIP may affect AP course and exam taking, college going, and subsequent college performance. I discuss these three margins in turn.

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. The gains to a student of taking and doing well on AP exams are also greater under the APIP. A principal agent multitask model (Holmstrom and Milgrom 1991) predicts that where good AP performance is more likely with higher teacher and student effort, both teachers and students will exert more effort to improve student AP output. Therefore, one would expect (A) an increase in teacher effort to recruit students to take AP courses, (B) an increase in teacher effort to improve the quality of their instruction, (C) an increase in student AP exam taking, (D) an increase in student effort to perform well in AP exams, and (E) an increase in AP course enrollment.

In a world with completely rational high schoolers, full information, no supply constraints, and perfectly functioning credit markets, the relatively small financial rewards for students of $100-$500 for taking and passing AP courses should little effect (as students will balance the *lifetime* benefits to taking AP courses against the immediate costs). However, the cash incentives associated with the APIP may produce a large effect if students are myopic, are 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, the APIP may improve the observable characteristics of affected students. These students may be more desirable candidates to college admission committees 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 directs of college attendance and therefore increasing enrollment.

While the APIP should increase college going, conditional on applying for college, it is not obvious that the APIP will increase college going overall since the APIP could potentially affect students' college application decisions. The decision to attend college is an investment under uncertainty. Students may not know how much they will enjoy college, or their likelihood of success at college. 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.[17] If students are pessimistic about their chances of success at college, then the APIP may lead them to adjust their expected costs/benefits of attending college so that they may be more likely to apply. Alternatively, if students are optimistic about their relative costs to college, then the APIP could reduce the likelihood that students apply to college. Recent findings by (Stinebrickner and Stinebrickner 2009) suggest the latter is much 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*.

In sum, the total effect will reflect a combination of the effect on students' college application behaviors and the effect on the likelihood of being admitted to college conditional on applying. While one would expect the APIP to increase the college going of college applicants, the total effect on college going is ambiguous in sign.

III.3 *Effect on College Performance*

[17] 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.

As discussed above, the APIP increases student exposure to more rigorous material. As such, the APIP should increase student knowledge, which in turn should be associated with improved academic outcomes. All else equal, if the APIP only affected students by increasing their exposure to AP courses, APIP exposure would be associated with unambiguously higher student achievement in college. I refer to this as the human capital mechanism.

However, there are reasons to expect that the APIP may not improve student college outcomes: First, there is evidence that test scores can be improved by simply increasing test taking effort (Braun and Kirsch 2008), having a good meal the day of the exam (Figlio and Winicki 2005) or gaining familiarity with the test format. As such, the improved qualifications of students such as having more AP passes, or higher SAT scores may not reflect actual 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 the monetary incentives, so that they may actually perform worse after incentives are removed than they would have if rewards were never offered (Kohn 1999). If this lowered intrinsic motivation phenomena applies to the APIP population, APIP affected students may actually perform *worse* in college than unaffected students. Third, a principal agent multitask model predicts that improved AP performance may come at the expense of other important skills if teachers teach-to-the-test or student neglect their non-AP courses. Fourth, the APIP may make students overly ambitious such that they apply to more difficult programs than they otherwise would and actually have worse outcomes as a result. That is, the APIP could lead to a sort of "mismatch"[18] between students and colleges that may ultimately lead to worse outcomes. This mismatch hypothesis is counter to the information story – such that affected students are *more* likely to make sub-optimal college going decisions.

In sum, while a simple human capital explanation suggests that the APIP would improve college outcomes, there are a variety of reasons why the APIP may have no effect on college outcomes and could *potentially* lead to worse college outcomes. As such, theory alone cannot tell us the sign or magnitude of the effect on college outcomes so that the total effect of the APIP on college outcomes remains an empirical question.

[18] The "mismatch" hypothesis states that students who would not ordinarily be admitted to selective schools may be inadequately prepared, such that they would fare better at schools better matched to their preparation (Summers 1970).

IV. The APIP Schools and The Data

Before turning to the regression data, to show how the APIP school differ from other schools in Texas, I present some basic statistics from the Common Core of Data from the National Centre for Education Statistics and the Texas Education Agency (TEA). These data are summarized in Table 1. Schools that were selected for the APIP were quite different from schools that have not yet been selected and may never be selected for the APIP. The APIP schools had average total enrollments during 2000 through 2005 of 1836 students – much larger than the average enrollment of 751 students for non-APIP schools in Texas. During 2000 and 2005, 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 same years only 25 percent of the students at APIP schools were white compared to 53 percent for non-APIP schools and about 10 percent of students were limited English proficient at APIP schools compared to less than 4 percent for non-APIP schools. Both groups of schools, however, have similar shares of economically disadvantaged students- reflecting the fact that Texas has both urban poor and rural poor.

The regression data used combines student records from every public and private tertiary institution in Texas[19] between 1999 and 2008 from the Texas Higher Education Coordinating Board with student-level high-school and middle-school data from the TEA. Advanced Placement examination data come from the College Board. The TEA provided the 8th grade and 10th grade statewide standardized test data for all Texas students between 1994 and 2007.[20] Texas law requires that all students take the state level achievement tests so that these test scores are not subject to selection bias.[21] These test score outcomes have been normalized and standardized to be mean zero and have a standard deviation of one for each test administration. For each student, I use the most recent administration of the test (that is, I throw out early attempts) so that mean test scores are above zero.[22] The final dataset contains the college outcomes, high-school and middle-school data of all students who were in 10th grade (the grade before exposure to APIP) between 1994 and 2007. Using the population of 10th graders allows me to account for attrition that may take place after 10th grade in 11th and 12th grade due to the

[19] 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 heath related institutions.
[20] TAKS (1994-2003) and TASP (2003-2007).
[21] 115 STAT. 1425 is No Child Left Behind
[22] 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

APIP.[23] Since college data span the years 1999 through 2008, college outcomes will only be available for students who were in 10th grade before 2006.

I present the pre and post APIP adoption summary statistics for the schools that will have adopted the APIP by 2008 (*note that schools adopt the APIP at different times so that the pre-adoption years differ across schools*). In the pre-adoption years, the likelihood that a 10th grader took an AP course while in high school was 0.1729, which increased to 0.25 in the post adoption period. There were similar increases in AP examination taking where 10th graders, on average, took 0.097 exams in the pre adoption period and took 0.1268 exams in the post adoption period. Looking at examination taking, about 5.5 percent of 10th graders took any AP exams in the pre-period compared to 6.8 in the post adoption period. Both 8th grade and 10th grade math and reading standardized scores were lower after adoption than before APIP adoption - suggesting some possible negative selection by ability into APIP schools.

In the pre-adoption years, about 42 percent of the 10 grade students in the sample attend any college and about 11 percent attend a four-year college. While 42 percent attend some college only 21 percent enroll in a second year of college. There are some important differences by student ethnicity discussed in section VI.2. It is important to note that the post years are *by definition* more recent years so that the outcomes defined for those who enroll in college appear to decline in the post years relative to the pre years. This reflects the fact that a student who graduated from high school in 2004 is much less likely to have entered her third year of or graduate from college by 2008 than a student who graduated from high school in 1997. As such, comparisons of the pre to the post adoption outcomes that don't take high school graduation year into account are meaningless. However, meaningful comparisons are made in a regression context and are presented in Section VI.

V. Empirical Strategy

Before presenting the identification strategy, in section IV.1 I discuss a few methodological concerns facing this and other similar studies and I present my proposed solutions. Specifically, I discuss how I deal with sample selection bias, bias due to controlling

[23] 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 APIP school 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.

for endogenous covariates, attrition bias, and endogenous treatment. In section IV.2, I present the identification strategy and discuss the source of plausibly exogenous variation.

V.1 Methodological Issues

Becasue this study aims to uncover the effect of the APIP on student college outcomes, it is important to compare the outcomes of students who were exposed to the APIP to the outcomes of similar students who were not exposed to the APIP. Much of the literature on college outcomes has looked at college outcomes of students with similar observable characteristics upon college matriculation. Since the APIP affects students in 11^{th} and 12^{th} grade, it may affect the observable characteristics of students upon matriculation by improving their test scores while still in high school so that one must compare students who were similar *before* exposure to the APIP. Many studies that have attempted to isolate a causal effect of AP exam scores on college outcomes by controlling for covariates such as SAT/ACT scores and senior year high school GPA (Geiser and Santelices 2004, Eimers 2003).[24] Since SAT/ACT scores and high school GPA could be affected by exposure to AP courses, these "covariates" that are determined after 10^{th} grade may be endogenous. To avoid such endogeneity, because the APIP affects students who are in 11^{th} and 12^{th} grade, I compare the outcomes of students with similar 10^{th} grade test scores from the same high-school and I *do not* control for potentially endogenous covariates such as SAT scores or senior-year high-school GPA.

Other important methodological issues are (a) the choice of population and (b) how the outcomes are defined. Students on the margin of going to college are likely to have lower preparation and motivation on average, such that these students may also be less likely to succeed in college upon matriculation. As such, if the APIP increases the number of students who attend college, the population of college students who were exposed to the APIP may have worse outcomes for reasons unrelated to the effect of the APIP because the APIP-exposed students will have a higher proportion of marginal college attendees. In short, using the sample of college students leads to sample selection bias. To isolate the effect of the APIP on college outcomes one cannot use a sample of college enrollees, but must compare the college outcomes of all potential college students. Basing the estimates on the population of prospective college students allows one to uncover the true causal effect, but it also introduces another methodological issue because students who do not enroll in college have no college outcomes.

[24] A notable exception is (Dougherty, Mellor and Jian 2006) who control for 8^{th} grade test scores.

One simple way to deal with missing outcomes for student who do not enroll in college is to impute values for those students. For many of the outcomes analyzed this is a natural solution. For example, a student who does not attend college is coded as not attending their sophomore year of college and not graduating from college. However, for outcomes like freshman year GPA, it is unclear what the value should be for those who do not attend college. The first approach used in this paper is to assume that anyone who does not enroll in college has a GPA of zero. In this case the modified GPA variable as [1] below where $I_{college=1}$ is equal to one for college attendees and zero otherwise.

[1] $$GPA = I_{college=1} \times (GPA \mid I_{college=1} = 1) + 0 .$$

Using the product rule, the expected change in GPA due to the APIP can be written as [2] below.

[2] $$\begin{aligned}\Delta E[GPA] \equiv \Delta[P(I_{college=1} = 1)] \times (GPA_0 \mid I_{college=1} = 1) + \\ P_0(I_{college=1} = 1) \times \Delta(GPA \mid I_{college=1} = 1)\end{aligned} .$$

In [2], P_0 is the college going rate for untreated students and GPA_0 is the mean GPA of untreated students who enter college. Equation [2] shows that changes in freshman year GPA will reflect an extensive margin effect (the effect of having a non-zero GPA due to attending college) represented by $\Delta[P(I_{college=1} = 1)] \times (GPA_0 \mid I_{college=1} = 1)$ and an intensive margin effect (due to improvements in GPA for students who would have gone to college even in the absence of the APIP) represented by $P_0(I_{college=1} = 1) \times \Delta(GPA \mid I_{college=1} = 1)$. Equation [2] makes explicit that changes on this GPA measure will not only reflect improvements in GPA for students who enroll in college, but also reflect improvement do to having more students enroll in college and earn a non-zero GPA. While this does not have a natural interpretation, it is a useful starting point and a helpful summary statistic for overall improvements in GPA.[25]

Because we are interested in both the effects overall and the effects on those students who would have enrolled in college without the APIP, I use two methods to uncover the effect conditional on college attendance. The first method is to impute GPAs for students who do not enroll in college, and then show the main results under different values of the imputed GPA. the estimates obtained over a reasonable range of imputed values will be informative of the effect conditional on enrolling in college. The second approach, proposed by (Angrist 1995), is to use

[25] In fact, if the APIP has no effect on college enrolment, one can easily uncover the effect conditional on college enrolment by dividing the estimate obtained using the full sample by the probability of attending college.

the sample of college enrollees, while controlling for the likelihood of attending college (estimated using the full sample).[26] The findings under both methods are similar.

The final methodological issue is how treatment is defined. Because students may enroll at APIP schools in 11th and 12th grade in order 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 as my main variable of interest 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. 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. However, using ITT as the variable of interest yields a reduced form estimate of the APIP effect; the policy relevant estimate.

V.2 Identification Strategy

The basic 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 – that is, compare the difference in outcomes across ITT cohorts and non-ITT cohorts from the same schools to the difference in outcomes across cohorts in 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 be better in observable and unobservable ways from students who attend schools that do not adopt the APIP. Also, by comparing cohorts, as opposed to comparing students within the same cohort,

[26] Trimming techniques and maximum likelihood approaches have been used by researchers (Angrist, Bettinger and Kremer 2006, Lee 2002). However where treatment is not binary and there are large vectors of fixed effects (as in this case) employing such methods pose significant implementation problems.

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 10^{th} grade test scores, I address the concern that the incoming preparation of students may have changed in APIP schools after adoption of the program.

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 the changes in comparison schools to be a credible counterfactual for what the APIP schools would have experienced in the absence of the APIP, the comparison schools must be similar to the APIP adopting schools in both observable and *unobservable* ways. Since APIP schools and non-APIP schools have very different observable characteristics, as shown in Table 1, using all other Texas high schools as the comparison group would be misguided. Due to a scarcity of donors, AP Strategies could not implement the APIP in all interested schools at the same time. This allows me to restrict the estimation sample to only those schools 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 two important benefits: (1) since APIP willing schools are observationally similar, they are likely to share common time shocks and (2) since all schools that agreed to adopt the APIP are similarly motivated and interested, restricting the sample in this way avoids comparing schools with motivated principals who want to adopt the APIP to schools with unmotivated principals who have no interest in the program. This sample restriction controls for school self-selection on time invariant *unobserved* characteristics, potentially allowing for a consistent estimate of the Average Treatment Effect on the Treated (ATT).

Within the sub-sample of APIP schools, identification relies on the assumption that the exact timing of APIP implementation is exogenous to other within-school *changes*. Since all willing schools had to wait for a donor to adopt the APIP, and timing of actual adoption relies on idiosyncratic donor preferences and availability, this assumption is plausible. However, there remains the concern that if donors selected schools based on the enthusiasm of school principals and administrators, or if some schools expressed interest before others, then the timing of adoption may not be orthogonal to *changes in* school characteristics. Since donor choices are not random, I cannot entirely rule this out. However, it is important to note that all regressions use within-school variation so that differences in time-constant school enthusiasm or motivation will not confound the results. Problems would only arise if expressing interest in the APIP, and thus adoption, were co-incident with *changes* in unobserved school enthusiasm or motivation. In section VI, I show that improvements only take place *after* APIP adoption and that the timing of when a school likely expresses interest in the APIP is *not* associated with improved outcomes - 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).

$$Y_{ich} = \beta_1 X_i + \beta_2 A_i + \sum_{k=1}^{4+} \mu_k I_{ITT\ year=k} + \theta_h + \theta_c + \varepsilon_{ich} \qquad [3]$$

In [3], Y_{ich} is the outcome of student *i* in graduating cohort *c*, from high school *h*. X_i is a *matrix* of student demographic characteristics such as race, gender, and free-lunch status while in high school. A_i is a vector of student achievement scores from 10th grade. $I_{ITT\ year=k}$ is an indicator variable denoting the ITT year, so that μ_1 is the effect of the APIP in its first intention to treat year and μ_k is the effect of the APIP in its kth intention to treat year. Specifically, to identify the effect of the APIP over time, 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 is a measure of how long the APIP has been in place. More specifically, 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). It is important to point out that all cohorts in ITT years greater than 1 have two years of exposure to the APIP while the first affected cohorts (ITT year=1) are only exposed to the program for one year. To control for differences in student attributes across high-schools, secular differences in performance over time, and differences in outcomes across cohorts, I include high-school fixed effects θ_h, cohort fixed effect θ_c. Standard errors are adjusted for clustering at the school level.[27]

VI. Main Results

[27] The regression analysis was conducted using SAS. The command for the fixed effects model with clustered standard errors (at the school level) used was "Proc Surveyreg".

Graphical Evidence: Before showing the regression results, it is helpful to present some visual evidence of a true causal APIP effect. Figure 2 shows the results of estimating a flexible version of equation [3], where I estimate effects for both pre-adoption years and post adoption years. For each outcome, I plot the estimated coefficients of adoption years -5 through 4 (the first year of adoption is year 0 in the figure).

As one can see, prior to adoption there is no trend in the number of AP exams passed, however in the first adoption year the number of exams passed jumps up and increases over time ─ clear visual evidence of an APIP effect on the number of AP exams passed that is not driven by underlying trends. Looking at freshman GPA and persistence to sophomore year, there is no evidence of an increasing trend in the pre-adoption years and there is an increase beginning in the first adoption year with a steady increase thereafter. To see if these increases are driven by increases in the number of college enrollees, as opposed to improvements in the achievement of those students who would have enrolled in college in absence of the APIP, I also show the effect *only* for students who enrolled in college (while controlling for the likelihood of enrolling in college). Persistence to freshman year conditional on enrolling in college and freshman year GPA conditional on enrolling in college both increase after APIP adoption but not before. It does appear that there may be some underlying trends in college enrollment. To test for trending formally, *for all outcomes*, 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 20 percent level. In contrast, 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 on all outcomes that is not driven by underlying trends.

The visual evidence suggests that the APIP increased the number of AP exams passed and increased college enrollment. Moreover, affected cohorts had higher freshman year GPAs, and were more likely to persist to sophomore year (overall and conditional on enrolling in college) ─ indicating improvements in college outcomes overall *and even for those who would have enrolled in college absent the APIP*.

Regression Results: Table 3 presents the regression results for AP participation, college enrollment, freshman year GPA and persistence to sophomore year of college. I analyze college graduation separately at the end of this section. For each outcome I report the coefficient on the first, second, third and fourth plus intention to treat year. Columns 1 and 2 show that the APIP is associated with a statistically insignificant 0.21 increase (about a 25 percent increase) in the number of AP courses taken by the fourth year, and a statistically significant 0.044 increase (about a 45% increase) in the number of AP examinations passed. Both these effects increase with greater APIP exposure, suggesting that the APIP had a large effect on AP course participation and AP examination achievement.

Columns 3 and 4 show the effect on the likelihood of matriculating in college and matriculating in a four-year college, respectively. While there is no effect on college matriculation in the first two years of the program, by the third year college matriculation increased by a statistically significant 0.034 (about an 8 percent increase). Column 4 shows that there was no effect on matriculating in a four-year college so that this increased college going is driven exclusively by increased two-year college attendance.

Column 5 shows the effect on freshman year GPA (on a four point scale). In this model, students who are not enrolled in college are assigned a GPA of zero, so that increases in this outcome reflect both increases in the likelihood of attending college and the effect of improved performance conditional on college attendance. As with college attendance, while there is no effect on college matriculation in the first two years of the program, by the fourth year of the APIP freshman GPAs increased by 0.09 grade points (significant at the 5 % level). Using the method proposed by (Angrist 1995), I estimate the likelihood of attending college using the full sample[28], then I estimate the main models only using the sample of college enrollees, *while controlling for the estimated likelihood of attending college* as a control function. This approach should yield the average effect of the APIP for those students who would have attended college absent the APIP. Results from this model are presented in Column 6. While there is no effect on freshman year GPA conditional on enrolling in college in the first two years of the program, by the fourth year of the APIP freshman GPAs increased by 0.066 grade points (significant at the 5 % level). The smaller effects conditional on enrolling in college suggest that while the APIP improves the freshman year GPAs of students who would have matriculated in college regardless of APIP exposure, some of the overall GPA effect was due to students who would not have attended college (and had an imputed GPA of zero) being more likely to enroll.[29]

To shed further light on how much of the APIP effect is driven by the extensive margin

[28] I estimate a probit model of attending college on all control variables and obtain an estimated propensity.
[29] It is easy to show that *if there were no effect on the extensive margin* the effect on the intensive margin would be the overall effect divided by the likelihood of attending college. This number must be larger than the overall effect.

(college entry) as opposed to the intensive margin (GPA conditional on enrolling in college), in Figure 3, I present the coefficient on the fourth year effect under different imputed GPA values for those who do not attend college. If one assumes that students who do not enroll in college would have received a GPA of 0 (an F average) then the fourth year effect would be 0.09 grade points. If one assumes that students who do not enroll in college would have received a GPA of 1 and 2 (a D and C average, respectively) then the fourth year effect would be about 0.06 and 0.035 grade points, respectively (both statistically significant at the 5 percent level). Even if one were to make the unrealistic assumption that students who do not enroll in college would have had on average the same GPA as those who did enroll (2.1 just above a C average) the fourth year effect would be about 0.027 grade points. In order for the estimated APIP effect to fall to zero would require that students who do not enroll in college would have GPAs above 3 (above a B average) ─ this is highly unlikely. The results in Figure 3 are consistent with the regression results and suggest that APIP exposure is associated with improved freshman year GPAs among those students who would have attended college without the APIP.

The last main outcome in Table 3 is persisting to sophomore year (that is, being enrolled as a sophomore in college). 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. Column 7 shows the results for the full sample and column 8 shows the results conditional on college enrollment (while controlling for the likelihood of enrollment). As with the other outcomes, while there is no effect on persistence in the first two years of the program, by the fourth year of the APIP persistence increased by a statistically significant 0.047 percentage points (about a 22 percent increase). Conditional on attending college this fourth year effect is a statistically significant 0.078 percentage points ─ indicating that much of the persistence effect is for students who would have enrolled in college irrespective of APIP exposure.

***Effect on different GPA margins*:** Because the freshman GPA is continuous one can look at the effect of the APIP on the likelihood of having a freshman GPA above any given value. Defining the outcome in this manner allows one to get a sense of whether increases in average freshman year GPA are due to improvements among students with high GPAs or are driven by improvements among students with low freshman year GPAs. Figure 4 plots the fourth year effect on having a freshman year GPA above various levels for all students and also only for those students who enrolled in college while controlling for the likelihood of college enrollment.

Based on all tenth graders, the APIP increases the likelihood of having a college freshman year GPA above 0.3 (an F+) by about 3.1 percentage points, the likelihood of having a college freshman year GPA above 1 (a D) by about 3.1 percentage points, and the likelihood of having a college freshman year GPA above 1.7 (a C-) by about 3.1 percentage points. The effect on higher GPA margins are small so that the APIP increases the likelihood of having a college freshman year GPA above 3 (an B) by about 0.5 percentage points (not statistically significant). The figure shows that the increases in average freshman year GPA are largely driven by improved freshman year GPAs for students with worse than a B- average, and are not driven by improvements among high GPA earners (those with GPAs above a B average).

The dashed line shows the same models conditional on attending college. Among college enrollees, the APIP effect on freshman year GPA is driven by students with GPAs between 1 and 2.3 (a D and C+ average, respectively). Based on college enrollees, the APIP increases the likelihood of having a college freshman year GPA above 1 (a D) by about 2.5 percentage points, and increases the likelihood of having a college freshman year GPA above 2.3 (a C+) by about 3 percentage points. Much like the unconditional model there are no APIP effects for GPAs above a B average. However, unlike the unconditional model there are no APIP effects for GPAs below a D average ─ implying that the increases in the likelihood of having GPAs above 0.3 and 0.7 were primarily due to students being more likely to enroll in college and have GPAs that are above zero. The similarities in the estimated effects for GPAs above 1.7 in the unconditional model and the conditional models is striking. Since the estimated effects conditional on enrolling would be larger if there were no extensive marginal contribution, it is clear that while improvements in GPA between 0 and 1 are due to the extensive margin, improvement in GPA between 1 and 2.7 are due to both the intensive and the extensive margins. Across both models, however, the APIP has no effect on the GPAs among students with better than a B average.

VI.1 Threats to Validity and Endogeneity Concerns

Section VI shows that APIP adoption is associated with statistically significant improvements in college outcomes for affected cohorts. While I am careful to make comparison across 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 may be the lingering worry that the findings do not reflect a causal relationship. The remaining endogeneity concerns are that the results could driven by (1) *changes* in school motivation, (2) student self-selection to treatment schools, (3) spurious correlation due to pre-existing underlying trends, or (4) students being more likely to enroll in Texas schools as a result of the APIP. In this section I will present evidence that the first three factors *do not* drive the results. I rule out the fourth concern in section VI.4 in the context of all the findings.

The timing of APIP adoption may be endogenous.

There is the concern that that schools that had an *increase* in motivation were more likely to apply to have the APIP implemented. If this were so, one would expect to see an improvement in outcomes before APIP adoption. 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 year prior to actual adoption. Figure 2 plots the estimated coefficients on each of five pre-treatment year indicator variables and each of five post-treatment year indicator variables. Visually, it is clear that none of the outcomes exhibit any significant trending prior to APIP adoption. This is supported by the fact that one cannot reject the null hypothesis that the coefficients on the pre-treatment year indicator variables differ from that of the year prior to APIP adoption are equal to zero at the 10 percent level. As a direct test of the "timing of interest" hypothesis, I regressed the outcomes on a an indicator that was equal to 1 two years prior to APIP adoption and thereafter and equal to 0 otherwise. This yielded very small coefficient estimates and *p*-values larger than 0.4 for all outcomes. This is consistent with assertions from AP Strategies officials that the timing of adoption is largely idiosyncratic, and is compelling evidence that APIP adoption was not endogenous to unobserved *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. If this were the case, one would expect to see incoming 8^{th} and 10^{th} grade test scores increase after APIP adoption. If there were positive selection driving the results, then the APIP should be associated with higher incoming test scores. To test this hypothesis, I estimate equation [3] using 10^{th} grade and 8^{th} grade test scores as the dependent variable, while controlling only for school fixed effects and year fixed effects. The results are in Table 4. Columns 1 through 4 show that the APIP is associated with lower incoming 8th and 10th grade test scores, so that any selection into APIP school is likely to be negative. Columns 5 through 8 show how observable demographics changes at APIP schools after adoption. One can see that in addition to having students with lower incoming test scores, APIP school became less low-income and less Hispanic after APIP adoption.

Since there may be negative selection into APIP schools, it is important to show that the results in Table 3 are not driven by controlling for this negative selection. To show that this is not the case, I estimate equation [3] without including any student controls (i.e. only including high-school and cohort fixed effects). The results are in presented in the top panel of Appendix Table 1. The estimated effects without controls are about 15 percent less than those in Table 3, supporting the assertion that the selection on observable characteristics imply a downward bias rather than an upward bias. In any case, while the coefficients obtained without controls are slightly smaller, they are well within the 90 percent confidence interval for the main estimates in Table 3, and tell the same basic story ─ showing that the estimated positive APIP effects are robust to omitting controls for negative selection into APIP schools.

There may be selection in unobserved dimensions

While the results indicate that selection on observables is negative one could worry that selection on unobservables is positive. To ensure that the results are not driven by selective migration on unobservable characteristics I estimate equation [3] 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 self-select). As such, 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. I also remove all students who attended middle schools that sent fewer than 300 students to any given APIP high school during the sample period. This should remove almost all potential for bias from student self-selection to treatment. The results of this empirical specification (middle panel of Appendix Table 1) are very similar to the results in Table 3, indicating that student self-election to high-schools does not drive the results.

APIP schools were already on a trajectory of improvement before adoption.

The visual evidence shows that 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 address pre-existing trends by including both a high-school specific intercept *and a linear time trend for each high school*. These results are presented in the lowest panel of Appendix Table 1. While the standard errors are larger, the point estimates are very similar to those in Table 3 with the exception of AP course taking, where the estimated fourth year effect is *larger* with the inclusion of a linear time trend for each high-school.

VI.2 Effects by Gender and Ethnicity.

In light of findings that cash incentives are associated with improved outcomes only for females, and a broader 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. To answer this question, I estimate equation [4] for males and females separately. Like (Jackson 2010) I find no evidence of differences by gender. Appendix Table 2 shows results for females, which are essentially the same as that for the full sample.

(Klopfenstein 2004) has documented that there are large and important differences in AP participation across ethnic groups both across school *and within schools*. To get a sense of differences across ethnic groups in AP outcomes and college outcomes in these data, I present summary statistics for key outcomes in Table 5 for white, black, and Hispanic students separately. About 17 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 18 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, whites have much higher AP participation and pass rates. About 27 percent of white 10th graders in the sample take any AP course, 10.6 percent take any AP exams, and the average number of exams passed is 0.113 (over ten times the mean number of exams passed for black students, and just under five times as many as Hispanic students). However, some of these differences may reflect differences in preparation as the average black and Hispanic 10th grader had math and reading test scores between 0.4 and 0.5 standard deviations lower than whites in the sample.

To test for differences in the APIP effect by ethnicity, I estimate equation [4] separately for white, black and Hispanic students in Tables 6,7 and 8, respectively. For whites students (Table 6), by the fourth year, there is a statistically significant 0.13 increase in the number of AP exams passed (a 115% increase). By the fourth year, white students were 3.4 percentage points more likely to enroll in college. This effect represents a 7 percent increase driven entirely by increases in two-year college going but is not statistically significant at the 10 percent level. By the fourth post adoption year whites have GPAs 0.099 grade points higher and they are 4 percentage points more likely to persist to sophomore year (a 15 percent increase). Conditioning on college enrollment, the GPA effects disappears, while the persistence effect increases to 5.9 percentage points.

The effects for black students are generally larger and more likely to be statistically significant than for white students. For black students (Table 7), by the fourth year, there is a statistically significant increase of 0.011 more AP courses and a statistically significant 0.012 increase in the number of AP exams passed (a 133% increase). By the third year, black students were 3.5 percentage point more likely to enroll in college (a 10 percent increase) and 1.9 percentage points more likely to attend a four-year college (a 19 percent increase). They also had GPAs 0.143 grade points higher and they were 6.2 percentage points more likely to persist to sophomore year (a 45 percent increase). Conditioning on college enrollment, the GPA effects fall slightly to 0.116 grade points and the persistence effect increases to 7.8 percentage points.

Hispanic students were effected differently from both white and black students. For Hispanic students (Table 8), by the fourth year there is a statistically significant 0.016 increase in the number of AP exams passed (a 76% increase). There is very little evidence that the APIP increases college going among Hispanic students. In addition to having a small extensive margin response (unlike black and white students), Hispanic students also experienced small GPA effects. By the fourth year, they had GPAs only 0.058 grade points higher (not statistically significant). Despite the small GPA effects, however, Hispanic students were 6.3 percentage points more likely to persist to sophomore year (a 50% increase). Conditioning on college enrollment the persistence effect increases to 12.7 percentage points.

While all groups passed more AP exams, the proportional increase was largest for blacks and whites than for Hispanics. The point estimates indicate that white and black students were about 3 percentage points more likely to attend college, while the effect for Hispanic students was only 1.2 percent and was not statistically significant ─ implying that much of the increased college going was driven by black and white students. One very important difference across groups was that blacks were more likely to enroll in four year college while whites and Hispanics were only more likely to attend two year colleges. Looking at outcomes while at college, blacks and white had large GPA effects while Hispanic students experienced small GPA effects. However, when looking at persistence, arguably a more important outcome, the marginal effects for black and Hispanic students were larger than those for whites, both in absolute terms and in relative terms. This implies that the APIP may narrow educational gaps across ethnic groups.

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. Given that black and Hispanic students (with similar incoming test scores) do not have the same response to the APIP, this is unlikely. However, to test for this, I estimated the main models separately for students with 10 grade scores in the top third (both math and reading scores were in the top third) the bottom third (both scores in the bottom third) and the middle group. While there are small differences by incoming academic preparation, these differences were unremarkable and do not explain the differences observed by ethnicity, so that these differences by ethnicity *do not* merely reflect differences in response by scholastic preparation.

VI.3 Evidence on College Graduation

While the results thus far show that the APIP increased college enrollment and increased persistence to sophomore year both overall and conditional on enrolling in college, even though showing an increase in years of education completed is very important, arguably the most important outcome is graduating from college. Analyzing college going in these data is feasible, but must be interpreted with a an important caveat. The majority of the APIP schools adopted the program after 2000, so that the likelihood of classifying students as non-graduates who would graduate in 2009 or 2010 is real concern. Because many students in affected cohort were still enrolled in college at the end of the sample period, the graduation outcomes are right censored. As such, while I do analyze the graduation outcomes, these results should be interpreted with some caution. It is worth noting that this censorship bias is likely to attenuate the estimates.

Table 9 presents the main specification where the outcome variables are graduating with any degree and graduating with a four-year degree. I also present the graduation results for Hispanic, white, and black students separately. As one can see in columns 1 and 5, the APIP does not appear to have any effect on graduating from college with a degree or graduating from college with a four year college degree, overall. However, the results broken up by ethnicity reveal some important differences. By the fourth year of APIP adoption Hispanic and Black students were 3.3 percentage points and 2.7 percentage points more likely to graduate from college with a degree, respectively. These increases represent a 50 percent increase for blacks and a 66 percent increase for Hispanic students. For white students, the point estimates are negative, but not statistically significant, so that there may be no effect on the graduation rate for white students. Looking at graduating from a four-year college, the results indicate that by the fourth year of the APIP Hispanic and Black students were 2.7 percentage points and 2.5 percentage points more likely to graduate from college with a degree, respectively. These increases represent a sizable 69 percent increase for blacks and a 83 percent increase for Hispanic students. There is no statistically significant effect for white students.

Given the magnitude of these improvements for black and Hispanic students, it is important to establish that these effects are not an artifact of pre-existing trends. To shed light on this issue, I plot the coefficients on four pre-adoption years and four post-adoption years for the different ethnic groups. I show effects on both persisting to junior year of college and graduating from college with a degree in Figure 5. There is a clear increase in persisting to junior year of college for all ethnic groups after APIP adoption that is larger for Hispanic and black students than for white students. It is also clear that there may be some pre-existing upward trend in persisting to junior year for black and white students. The F-tests on pre treatment years yield *p*-values 0.58, 0.86, and 0.09 for white, Hispanic and black students respectively, suggesting some trending for black students.

Looking at graduating with any degree (right panel of Figure 5), while there is no evidence of trends for Hispanic students, there is an upward trend in graduating from college before APIP adoption, and evidence of a downward trend in graduating from college before APIP adoption for white students. The F-tests on pre-treatment years yield *p*-values of 0.12, 0.71, and 0.014 for White, Hispanic and black students, respectively — suggesting an upward trend for black students and a downward trend for white students. To assess the importance of these trends, I estimated models that include year trends for each school. Such models yield very similar results to those in Table 9 for whites and Hispanics, but reduces the estimated effect on the graduation rates of black students by about two thirds.[30] As such, the positive effect on college graduation for Hispanic students is strongly supported by the data, while I take the positive effect on college graduation for Black students as suggestive. While I cannot rule out the possibility that the improved graduation results for blacks are the result of pre-existing trends, given the large increases in college going, sophomore year persistence and junior year persistence, this is unlikely.

VI.4 Could Decreased Out-of-State College-Going Drive the Results?

Given that the APIP is associated with increased college going and improved outcomes while in college, one may wonder whether the results could be driven by high-achieving students being less likely to attend college outside of Texas (and therefore more likely to enroll in college in Texas) as a result of the APIP. Because I only observe students who attend college in Texas, such behaviors would lead me to interpret a switching of college going from out of state to in-state as an increase in overall college going. While this is a legitimate concern, I can rule out that this is important for the findings. Consider the following pieces of evidence.

(1) Students who attend college out-of-state overwhelmingly attend four-year colleges.[31] Given that there were no increases in four-year college enrollment, a "switching story" would require that students who would have attended four year colleges out-of-state decide to attend community colleges in Texas as a result of the APIP. This scenario is implausible.

(2) Among resource poor schools with low shares of white students (most APIP school fall into this category) about 2 percent of college bound seniors attend college outside Texas (Tienda and Niu 2006). Roughly two-thirds of 10th graders in the APIP schools graduate from high-school so this equates to *at most* 1.2 percent of 10th graders. Even if *all* out-of-state students moved in-state after APIP adoption, this could only increase college enrollment and persistence by 1.2 percentage points ─ this is much smaller than the estimated effects.

[30] Note that in models on the subsamples by ethnicity, controlling for school by year trends does not affect any of the other outcomes.

[31] This statement is based on an analysis of the Texas Higher Education Opportunity Project survey data of high school seniors in Texas.

(3) For all groups the increases in persistence are larger than the increases in enrollment. Among Hispanics, sophomore year persistence is 6.3 percentage points higher after four years of adoption while that for going to college is a statistically insignificant 1.2 percentage points. Also, while the percentage of Hispanic students graduating with a Bachelors' degree increased by 2.7 percentage points after four years of adoption, there was no increase in four year-college enrollment. Even under the highly implausible scenario that all the increased enrollment were due to shifting from out of state, it would be impossible for the persistence and graduation effects to be larger than the college enrollment effect unless there were a sizable causal APIP effect.

(4) The results that condition on college enrollment show positive APIP effects.

(5) As figure 4 illustrates, those students who were induced to attend college were those who were at the bottom of the achievement distribution (as one would expect) rather than at the top.

In sum, the prevalence of out-of-state college going among the students at APIP schools is too low to drive the estimated results, and *several* empirical patterns are inconsistent with the results being the result of increased enrollment from out of state ─ compelling evidence that students moving in-state due to the APIP does not affect the results in any meaningful way.

VI.5 Anecdotal Evidence of the Mechanisms

Given that the improved college outcomes appear to track well with increase AP examination participation, one may wonder what drives this increased AP participation. The large increases in AP participation are difficult to reconcile with the standard full-information full-rationality models of the schooling decision, and are much more consistent with there being sub-optimal effort on the part of teacher or students. While I am unable to test the mechanisms behind the APIP effect, I did obtain anecdotal evidence on why the APIP may have been effective. Evidence from discussions with guidance counselors at three different APIP high schools in Dallas strongly suggests 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. There is also evidence that certain barriers to taking AP courses were removed; at one high school, there used to be a minimum class rank that a student had to have in order to take AP courses, but after the APIP was adopted any interested student was allowed to take these courses. All guidance counselors mentioned a shift in student and teacher attitudes toward AP courses. AP courses are now considered difficult courses that anyone can take, as opposed to being available only for the very brightest of students (one AP teacher noted that she now has to turn students away). For example, one AP English teacher who had 11 students in 1995 and 110 students in 2003. Counselors credit the large increases in AP participation to student information, increased access through teacher encouragement, and increased teacher and guidance counselor recommendations. The financial incentives to students and teachers may have been responsible for the increased student and teacher effort in AP courses, but these aspects of the program were downplayed by the counselors.[32]

VII Conclusions

Using a carefully selected of group of comparison schools within which APIP adoption is likely exogenous, I find that students who were affected by the APIP were more likely to matriculate in college in Texas. Affected students were more likely to have higher freshman year GPAs, and there were improvements in college GPA even for those students who would have enrolled in college absent the APIP. The improvement in freshman year GPA, conditional on college enrollment was driven by improvements for students who would have had freshman year GPAs between a 1 and a 2.3 (between a D and a C+ average). The second measure of success at college is persistence beyond the first year. Students of all ethnic backgrounds and genders were more likely to persist to their sophomore year of college both overall and conditional on college enrollment ─ indicating that the APIP improved the overall educational attainment of affected students. Consistent with the results being the result of the APIP treatment, these positive effects

[32] Jackson (2010) finds that (1) APIP school that paid higher rewards did not have better outcomes than schools that paid less per passing AP exam, and (2) AP course enrollment increased in all AP subjects after APIP adoption even if rewards were not provided for all AP subjects. This is consistent with the claims of guidance counselors that the incentives *per se* may not be the driving force behind the APIP effects. However, because the level of incentives and the subjects rewarded are not exogenous, this evidence is suggestive and should be interpreted with great caution.

are only present after the second year of the APIP, after which the affected cohorts are exposed to the APIP during both 11th and 12th grade. I present empirical tests showing that the results are robust to including explicit controls for student self-selection and controlling for school specific time trends. I show that any selection bias is likely to attenuate the estimates, and I show that changes in out-of-state college going could not meaningfully affect the results.

While there are no significant differences by gender, the results indicate that the APIP increased college attendance for students from all ethnic groups, but led to larger improvements in student performance conditional on college attendance for black and Hispanic students than white students. In fact, the increase in the likelihood of persisting to sophomore year were equal to half the size of the black-white and Hispanic-white college persistence gaps.

Looking to college graduation, there is little evidence of an increase overall. However, when broken up by ethnicity, there is strong evidence of increased college graduation (from both two year and four year colleges) for Hispanic students and suggestive evidence of increased college graduation (from both two year and four year colleges) for black students. Because the graduation outcome is truncated (that is, students who are still in college are coded as not graduating) these results may understate the true effect of the APIP in college graduation. Taken at face value, the effects sizes on college graduation for Hispanic and black students are between one quarter and one third of the black-white and Hispanic-white gaps in college degree attainment in these data. This implies that programs like the APIP may be rather 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 sizably increased AP examination taking. Across all specifications and models, the APIP effect increases over time. This likely reflects some learning-by-doing effects and the fact that any curricular changes and early emphasis on pre-AP material would not have affected cohorts until a few years after the program had been adopted. Since the APIP combines student and teacher incentives with increased teacher training and curricular improvements, the results do not speak directly to the *isolated* effects of teacher performance pay, student incentives or teacher training. However, the finding that the program confers enduring benefits on students when extrinsic motivators are no longer provided is important for the literature on students 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 scheme. Also, the results present compelling evidence that the combination of these factors has important and sizable benefits.

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 $200*0.07=$14. Assuming a student is affected for two year this comes to $28 per 10th grader. By the fourth year of implementation, the APIP increases the likelihood of attending college for more than one year by 4.7 percentage points. Under the conservative assumption that those students who are more likely to persist to sophomore year only attend college for one additional year, this would lead to an increase in the average overall years of schooling of 0.047 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 $600 per 10th grade student. For this program to not be cost effective would require that the present discounted value of the lifetime benefits of an additional year of education be less than $600. This conservative estimate of the cost is orders of magnitude smaller than estimates of the benefits to an additional year of education, so this program is likely a worthwhile investment.[33]

Given that the large increases in AP participation are inconsistent with a standard Becker-Rosen model of schooling (implying 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, an reduced barriers to taking AP exams, it is not surprising that the economic returns to the APIP 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. While providing cash incentives

[33] 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.

for students and teachers may be one way to accomplish this, it may not be the only way. Given the large benefits associated with the APIP, further research is needed to deepen our understanding of how the program works, as such research may shed some much needed light on how one may improve the long-run academic outcomes of students.

Overall, the findings suggest that providing monetary incentives to both students and teachers to promote increased participation and improved performance in rigorous courses can lead to meaningfully improved student outcomes. The fact that the positive effects were larger for ethnic minority students suggests that programs that lead to increased rigor in high school 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." *Amercian 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, no. 1 (2009): 136-163.

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

Belley, 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.

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, 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.

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 fromSchool 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. 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.

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.

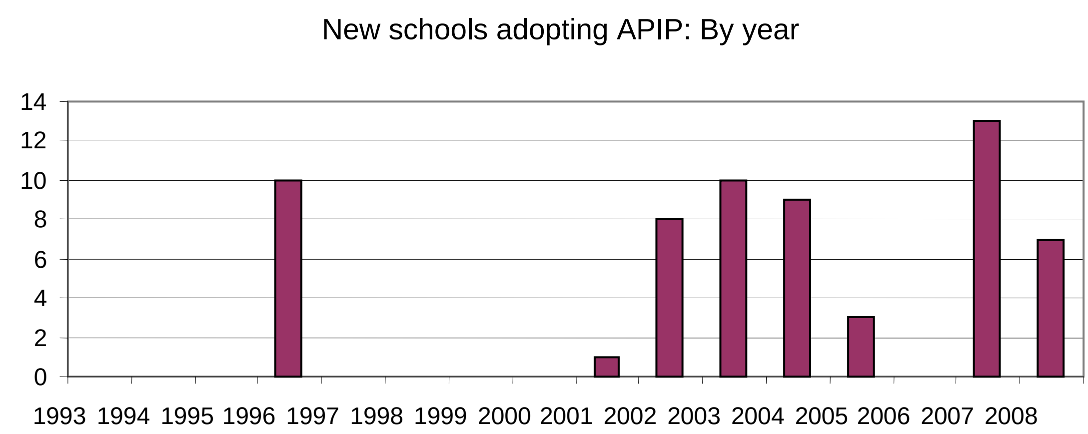
Seftor, Neil S., Arif Mamun, and Allen Schirm. "The Impacts of Regular Upward Bound on Postsecondary Outcomes 7-9 Years After Scheduled High School Graduation." Mathematica Report, 2009.

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.

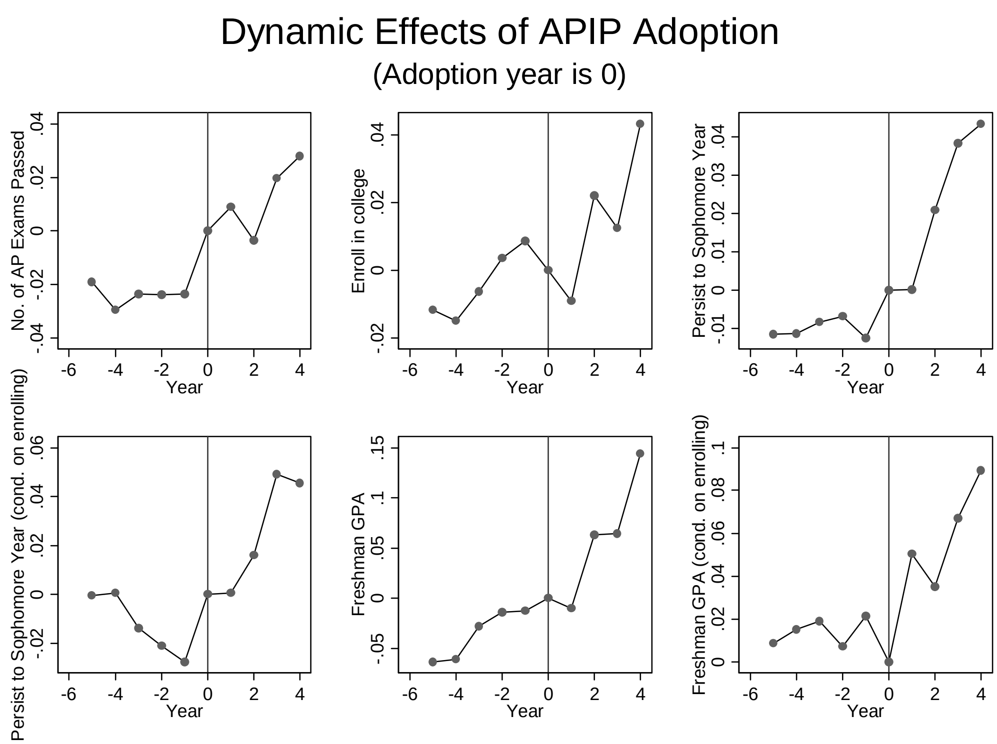
Tienda, Marta, and Sunny Xinchun Niu. "Flagships, Feeders, and the Texas Top 10% Law: A Test of the "Brain Drain" Hypothesis." *The Journal of Higher Education* 77, no. 4 (2006).

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



List of APIP districts: Abilene ISD, Amarillo ISD, Austin ISD (2008), Burkburnett ISD, Carrollton-farmers Branch ISD, City view ISD, Dallas ISD, Denison ISD, Galveston ISD, Houston ISD (2007), Pasadena ISD (2007), Pflugerville ISD, Tyler ISD, Wichita Falls ISD, Ysleta ISD (2007).

Figure 1: *New APIP schools by year.*



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.*

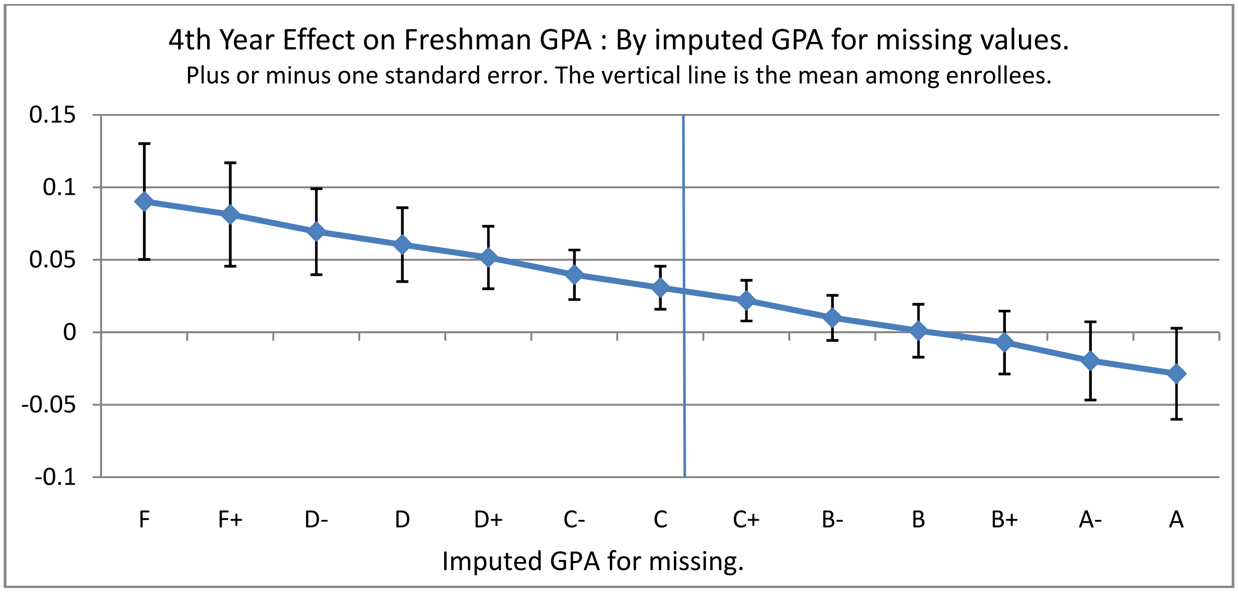


Figure 3: *Sensitivity to Imputed GPA.*

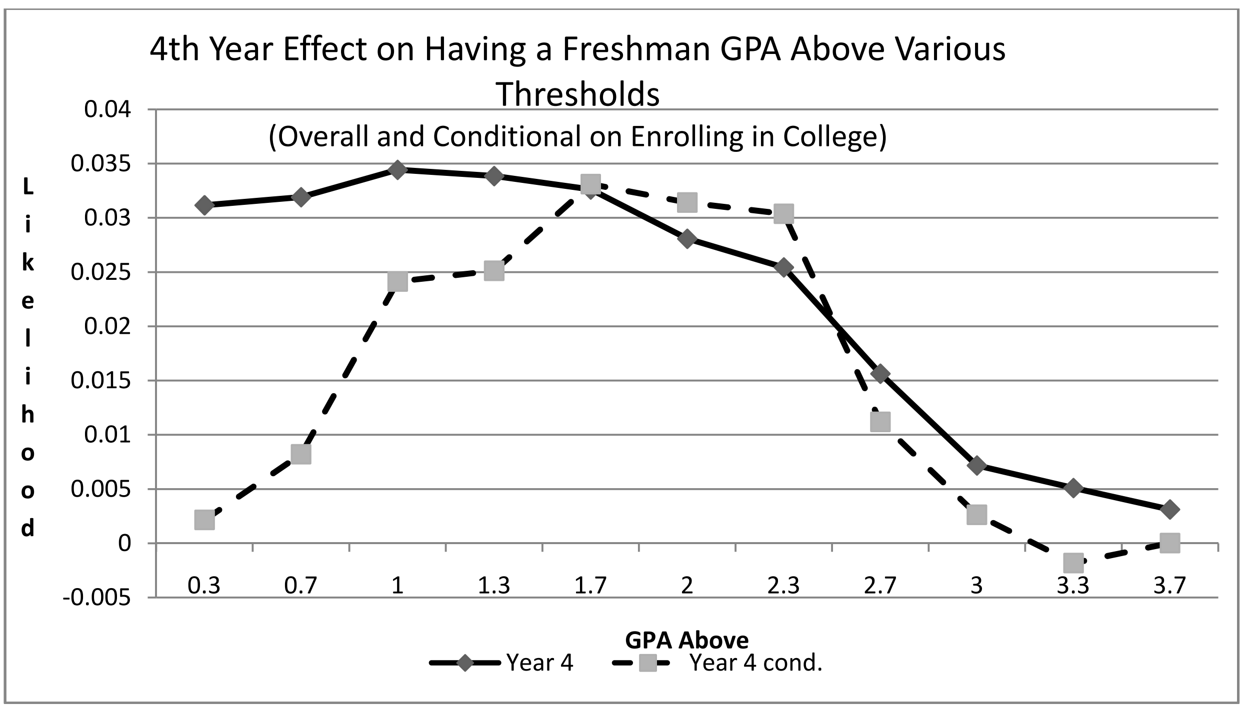


Figure 4: *Effects at Different GPA Margins.*

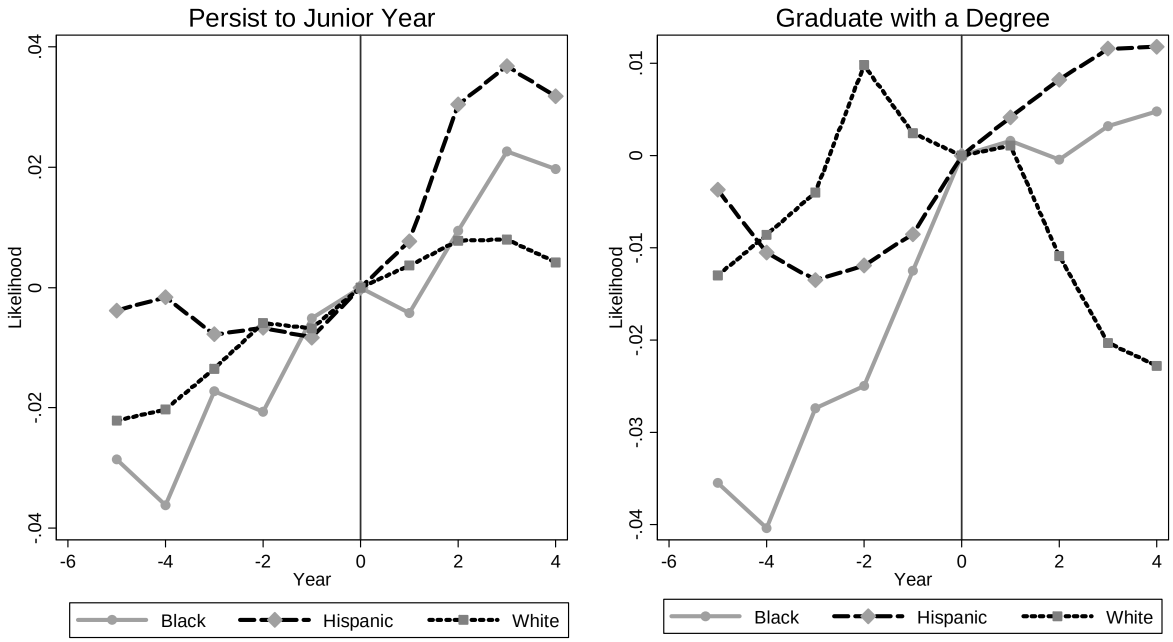


Figure 5: *Effect of the APIP on persisting to Junior Year and Graduating by Ethnicity*

Table 1

SUMMARY STATISTICS FOR DEMOGRAPHICS FOR APIP SCHOOLS AND OTHER COMPARISON GROUPS

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

Standard deviations in parentheses.

Table 2

Summary Statistics of the APIP schools

Variable	Not Adopted Mean	Not Adopted Std Dev	Adopted Mean	Adopted Std Dev
Take any AP course	0.1729	(0.378)	0.2521	(0.434)
AP courses taken	0.8258	(2.388)	1.3614	(3.155)
AP courses passed	0.7806	(2.323)	1.2895	(3.088)
Take any AP exam	0.0549	(0.228)	0.0682	(0.252)
AP exams taken	0.0974	(0.506)	0.1268	(0.598)
AP exams passed	0.0473	(0.342)	0.0535	(0.366)
Math score 10th Grade	-0.0907	(1.0)	-0.0806	(0.959)
Reading score 10th Grade	-0.0882	(1.013)	-0.0662	(0.987)
Math score 8th Grade	0.0568	(0.923)	0.0109	(0.954)
Reading score 8th Grade	0.0541	(0.919)	0.0112	(0.96)
Attend any college	0.423	(0.494)	0.3209	(0.467)
Attend a 4-year college	0.1101	(0.313)	0.073	(0.26)
Freshman year GPA	0.8956	(1.308)	0.6725	(1.22)
Earn an Associate's degree	0.0324	(0.177)	0.0131	(0.114)
Earn a Bachelors degree	0.0887	(0.284)	0.0295	(0.169)
White	0.3087	(0.462)	0.2553	(0.436)
Black	0.2047	(0.404)	0.2702	(0.444)
Hispanic	0.4445	(0.497)	0.4285	(0.495)
Asian	0.0346	(0.183)	0.0362	(0.187)
Native American	0.0035	(0.059)	0.0039	(0.062)
Female	0.5028	(0.5)	0.5091	(0.5)
Free or reduced lunch	0.3871	(0.588)	0.4642	(0.578)
Limited English	0.1126	(0.339)	0.1392	(0.356)
Persist at least one semester	0.361	(0.48)	0.2654	(0.442)
Attend college sophomore year	0.2121	(0.409)	0.134	(0.341)
Attend college junior year	0.1436	(0.351)	0.0733	(0.261)
Attend college senior year	0.1124	(0.316)	0.0462	(0.21)
Count	155753		138535	

Table 3

EFFECTS OVER TIME

SAMPLE IS ALL 10TH GRADERS WHO ATTENDED APIP SCHOOLS BETWEEN 1994 AND 2007

	1	2	3	4	5	6	7	8
	AP courses Taken	AP Exams Passed	Attend College	Attend 4 Year College	Freshman GPA	Freshman GPA Cond.	Persist to Sophomore	Persist to Sophomore Cond.
Mean of DV	1.08	0.05	0.375	0.093	0.7906	2.108	0.175	0.467
ITT years= 1	0.026	**0.031**	0.007	0.003	0.033	-0.003	0.015	0.013
	(0.123)	**(0.008)**	(0.014)	(0.008)	(0.036)	(0.02)	(0.007)	(0.011)
ITT years= 2	0.155	**0.042**	0.005	-0.003	0.026	0.048	0.015	0.012
	(0.126)	**(0.008)**	(0.019)	(0.009)	(0.043)	(0.024)	(0.01)	(0.014)
ITT years= 3	0.065	**0.029**	**0.034**	0.002	**0.091**	0.033	**0.034**	**0.031**
	(0.141)	**(0.009)**	**(0.015)**	(0.009)	**(0.035)**	(0.022)	**(0.01)**	**(0.014)**
ITT years= 4+	0.210	**0.044**	**0.030**	-0.002	**0.090**	**0.066**	**0.047**	**0.078**
	(0.145)	**(0.011)**	**(0.016)**	(0.012)	**(0.04)**	**(0.023)**	**(0.016)**	**(0.019)**
Obs	294288	294288	294288	294288	294288	110329	294288	110329
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

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

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

Table 4

SAMPLE IS ALL 10TH GRADERS WHO ATTENDED APIP SCHOOLS BETWEEN 1994 AND 2007

EFFECT OF APIP ON INCOMING TEST SCORES AND DEMOGRAPHICS

	1	2	3	4	5	6	7	8
	Grade 10 Math score	Grade 10 Reading score	Grade 8 Math score	Grade 8 Reading score	LEP	Low Income	Black	Hispanic
ITT years= 1	-0.029	0.004	-0.032	-0.007	0.012	-0.020	0.007	-0.014
	(0.024)	(0.021)	(0.022)	(0.018)	(0.008)	(0.015)	(0.01)	(0.011)
ITT years= 2	**-0.065**	**-0.045**	-0.053	-0.043	0.007	**-0.045**	0.003	**-0.024**
	(0.032)	**(0.026)**	(0.028)	(0.023)	(0.012)	**(0.021)**	(0.01)	**(0.013)**
ITT years= 3	**-0.075**	-0.037	0.006	0.023	0.010	**-0.053**	0.001	**-0.039**
	(0.037)	(0.033)	(0.041)	(0.035)	(0.016)	**(0.024)**	(0.015)	**(0.017)**
ITT years= 4+	**-0.056**	**-0.055**	-0.056	-0.039	0.019	**-0.067**	0.012	**-0.046**
	(0.035)	**(0.031)**	(0.045)	(0.042)	(0.02)	**(0.03)**	(0.017)	**(0.02)**
Obs	294288	294288	218669	218669	294288	294288	294288	294288
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust t-statistics in parenthesis are adjusted for clustering at the school level.

Table 5

Summary Statistics by Ethnicity

	Black		Hispanic		White	
	Parameter	SE	Parameter	SE	Parameter	SE
Take any AP course	0.172	(0.378)	0.180	(0.384)	0.270	(0.444)
AP courses taken	0.830	(2.345)	0.795	(2.263)	1.547	(3.411)
AP courses passed	0.780	(2.276)	0.729	(2.168)	1.498	(3.364)
Take any AP exam	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)
Math score 10th Grade	-0.339	(0.974)	-0.199	(0.952)	0.247	(0.93)
Reading score 10th Grade	-0.247	(1.045)	-0.199	(0.997)	0.233	(0.897)
Math score 8th Grade	-0.145	(0.9)	-0.100	(0.98)	0.360	(0.804)
Reading score 8th Grade	-0.095	(0.917)	-0.124	(0.989)	0.368	(0.781)
Attend any college	0.352	(0.478)	0.296	(0.457)	0.497	(0.5)
Attend a 4-year college	0.100	(0.3)	0.054	(0.227)	0.139	(0.346)
Freshman year GPA	0.617	(1.105)	0.593	(1.139)	1.176	(1.448)
Earn a any degree	0.053	(0.223)	0.050	(0.217)	0.147	(0.354)
Earn a Bachelors degree	0.039	(0.193)	0.030	(0.171)	0.115	(0.319)
Female	0.521	(0.5)	0.505	(0.5)	0.499	(0.5)
Free or reduced lunch	0.482	(0.572)	0.579	(0.646)	0.147	(0.355)
Limited English	0.009	(0.087)	0.259	(0.483)	0.006	(0.082)
Persist at least one	0.279	(0.448)	0.242	(0.428)	0.442	(0.497)
Attend college sophomore	0.139	(0.345)	0.127	(0.333)	0.259	(0.438)
Attend college junior year	0.079	(0.269)	0.068	(0.251)	0.186	(0.389)
Attend college senior year	0.055	(0.227)	0.046	(0.209)	0.144	(0.351)
Count	69315		128598		83441	

Table 6

SAMPLE: WHITE 10TH GRADERS WHO ATTENDED APIP SCHOOLS

	1	2	3	4	5	6	7	8
	AP courses Taken	AP Exams Passed	Attend College	Attend 4 Year College	Freshman GPA	Freshman GPA Cond.	Persist to Sophomore	Persist to Sophomore Cond.
Mean of DV	1.547	0.113	0.497	0.139	1.176	2.36	0.259	0.52
ITT years= 1	-0.354	**0.052**	-0.001	0.014	0.033	0.001	0.012	0.007
	(0.326)	**(0.019)**	(0.013)	(0.011)	(0.035)	(0.029)	(0.008)	(0.012)
ITT years= 2	-0.272	**0.067**	0.005	0.002	**0.058**	0.045	0.018	0.016
	(0.343)	**(0.019)**	(0.014)	(0.014)	**(0.037)**	(0.037)	(0.012)	(0.02)
ITT years= 3	-0.239	**0.054**	0.026	0.004	**0.089**	0.005	**0.031**	**0.032**
	(0.362)	**(0.018)**	(0.018)	(0.013)	**(0.042)**	(0.035)	**(0.012)**	**(0.018)**
ITT years= 4+	0.325	**0.130**	0.034	0.001	**0.099**	0.031	**0.040**	**0.059**
	(0.343)	**(0.029)**	(0.022)	(0.017)	**(0.053)**	(0.034)	**(0.019)**	**(0.023)**
Obs	83441	83441	83441	83441	83441	41478	83441	41478
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school level
All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 7

	Sample: Black 10th graders who attended APIP schools							
	1	2	3	4	5	6	7	8
	AP courses Taken	AP Exams Passed	Attend College	Attend 4 Year College	Freshman GPA	Freshman GPA Cond.	Persist to Sophomore	Persist to Sophomore Cond.
Mean of DV	0.83	0.009	0.352	0.1	0.617	1.752	0.139	0.395
ITT years= 1	**0.006**	**0.006**	0.002	-0.007	0.023	0.020	0.022	0.037
	(0.002)	**(0.002)**	(0.013)	(0.007)	(0.029)	(0.028)	(0.011)	(0.019)
ITT years= 2	**0.008**	**0.009**	0.007	-0.003	0.039	**0.036**	**0.018**	-0.003
	(0.003)	**(0.003)**	(0.014)	(0.007)	(0.029)	**(0.045)**	**(0.011)**	(0.021)
ITT years= 3	**0.006**	**0.007**	**0.035**	**0.016**	**0.126**	**0.062**	**0.038**	**0.019**
	(0.003)	**(0.003)**	**(0.013)**	**(0.01)**	**(0.03)**	**(0.036)**	**(0.013)**	**(0.02)**
ITT years= 4+	**0.011**	**0.012**	**0.029**	**0.019**	**0.143**	**0.116**	**0.062**	**0.078**
	(0.003)	**(0.003)**	**(0.016)**	**(0.01)**	**(0.039)**	**(0.043)**	**(0.019)**	**(0.027)**
Obs	69315	69315	69315	69315	69315	24378	69315	24378
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school level
All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 8

	Sample: Hispanic 10th graders who attended APIP schools							
	1	2	3	4	5	6	7	8
	AP courses Taken	AP Exams Passed	Attend College	Attend 4 Year College	Freshman GPA	Freshman GPA Cond.	Persist to Sophomore	Persist to Sophomore Cond.
Mean of DV	0.795	0.021	0.296	0.054	0.593	2.00	0.127	0.429
ITT years= 1	-0.090	**0.009**	-0.017	-0.019	-0.028	-0.034	0.006	0.013
	(0.089)	**(0.003)**	(0.021)	(0.01)	(0.049)	(0.035)	(0.011)	(0.018)
ITT years= 2	0.025	**0.016**	-0.040	-0.025	-0.088	0.032	0.005	0.031
	(0.098)	**(0.004)**	(0.03)	(0.013)	(0.063)	(0.045)	(0.013)	(0.029)
ITT years= 3	-0.124	**0.017**	0.008	-0.008	0.022	0.043	**0.034**	**0.054**
	(0.098)	**(0.006)**	(0.021)	(0.01)	(0.046)	(0.036)	**(0.01)**	**(0.024)**
ITT years= 4+	-0.043	**0.016**	0.012	-0.005	0.058	0.055	**0.063**	**0.127**
	(0.1)	**(0.006)**	(0.023)	(0.01)	(0.054)	(0.041)	**(0.014)**	**(0.028)**
Obs	128598	128598	128598	128598	128598	38103	128598	38103
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school level
All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Table 9

	Graduate with any Degree				Graduate with Bachelors Degree			
	1	2	3	4	5	6	7	8
	All	Hispanic	Black	White	All	Hispanic	Black	White
ITT years= 1	0.002	0.007	**0.026**	-0.002	0.005	**0.007**	**0.029**	0.003
	(0.006)	(0.004)	**(0.012)**	(0.009)	(0.006)	**(0.004)**	**(0.011)**	(0.008)
ITT years= 2	0.002	0.011	**0.029**	-0.003	0.005	**0.012**	**0.031**	0.004
	(0.01)	(0.007)	**(0.015)**	(0.013)	(0.009)	**(0.006)**	**(0.013)**	(0.013)
ITT years= 3	-0.002	**0.026**	**0.031**	-0.018	-0.003	**0.019**	**0.028**	-0.011
	(0.016)	**(0.011)**	**(0.015)**	(0.019)	(0.015)	**(0.01)**	**(0.013)**	(0.02)
ITT years= 4+	-0.010	**0.033**	**0.027**	-0.034	-0.013	**0.027**	**0.025**	-0.030
	(0.02)	**(0.013)**	**(0.014)**	(0.025)	(0.018)	**(0.012)**	**(0.011)**	(0.025)
Obs	294288	128598	69315	83441	294288	128598	69315	83441
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school level
All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.

Appendix

Appendix Table 1

ROBUSTNESS CHECKS

	1	2	3	4	5	6
	AP courses Taken	AP Exams Passed	Attend college	Attend 4 Year College	Freshman GPA	Persist to Sophomore
			Main Model With No Student Demographic Controls			
ITT years= 1	-0.008	**0.031**	0.005	0.002	0.030	0.013
	(0.123)	**(0.008)**	(0.016)	(0.007)	(0.038)	(0.008)
ITT years= 2	0.116	**0.041**	-0.003	-0.007	0.000	0.009
	(0.124)	**(0.008)**	(0.022)	(0.01)	(0.048)	(0.011)
ITT years= 3	0.019	**0.028**	**0.027**	0.000	**0.075**	**0.028**
	(0.145)	**(0.009)**	**(0.017)**	(0.01)	**(0.039)**	**(0.011)**
ITT years= 4+	0.188	**0.043**	**0.021**	-0.005	**0.072**	**0.039**
	(0.157)	**(0.011)**	**(0.017)**	(0.012)	**(0.042)**	**(0.016)**
			With Middle School by High School Pair Fixed Effects			
ITT years= 1	0.113	**0.039**	-0.006	0.001	0.001	0.008
	(0.129)	**(0.011)**	(0.017)	(0.01)	(0.044)	(0.01)
ITT years= 2	0.319	**0.047**	-0.006	-0.003	0.001	0.013
	(0.127)	**(0.011)**	(0.023)	(0.011)	(0.051)	(0.012)
ITT years= 3	0.109	**0.040**	**0.023**	0.006	**0.079**	**0.034**
	(0.148)	**(0.012)**	**(0.015)**	(0.01)	**(0.034)**	**(0.011)**
ITT years= 4+	**0.214**	**0.059**	**0.027**	-0.001	**0.079**	**0.048**
	(0.136)	**(0.017)**	**(0.017)**	(0.016)	**(0.041)**	**(0.019)**
			With School Specific Trends and Intercepts			
ITT years= 1	**0.184**	**0.034**	-0.004	0.006	0.016	0.013
	(0.143)	**(0.011)**	(0.018)	(0.011)	(0.047)	(0.009)
ITT years= 2	**0.405**	**0.045**	-0.004	0.000	0.016	0.015
	(0.14)	**(0.009)**	(0.024)	(0.013)	(0.059)	(0.011)
ITT years= 3	**0.382**	**0.041**	0.018	0.002	0.064	**0.028**
	(0.143)	**(0.011)**	(0.022)	(0.012)	(0.054)	**(0.011)**
ITT years= 4+	**0.581**	**0.066**	0.014	-0.003	0.070	**0.045**
	(0.19)	**(0.017)**	(0.024)	(0.014)	(0.06)	**(0.014)**

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school 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

	SAMPLE: FEMALE 10TH GRADERS WHO ATTENDED APIP SCHOOLS							
	AP courses Taken	AP Exams Passed	Attend College	Attend 4 Year College	Freshman GPA	Freshman GPA Cond.	Persist to Sophomore	Persist to Sophomore Cond.
ITT years= 1	-0.003	**0.075**	0.003	0.001	0.027	0.016	0.013	0.017
	(0.124)	**(0.013)**	(0.015)	(0.007)	(0.04)	(0.024)	(0.007)	(0.012)
ITT years= 2	0.063	**0.077**	0.005	-0.001	0.032	**0.057**	0.014	0.015
	(0.142)	**(0.017)**	(0.02)	(0.01)	(0.047)	**(0.028)**	(0.01)	(0.015)
ITT years= 3	0.014	**0.072**	**0.036**	0.006	**0.107**	**0.058**	**0.037**	**0.037**
	(0.151)	**(0.02)**	**(0.015)**	(0.01)	**(0.038)**	**(0.024)**	**(0.01)**	**(0.015)**
ITT years= 4+	**0.220**	**0.106**	**0.027**	0.001	**0.105**	**0.082**	**0.057**	**0.088**
	(0.152)	**(0.024)**	**(0.018)**	(0.013)	**(0.043)**	**(0.025)**	**(0.016)**	**(0.02)**
Obs	148842	148842	148842	148842	148842	60362	148842	60362
Year FX	YES	YES	YES	YES	YES	YES	YES	YES
School FX	YES	YES	YES	YES	YES	YES	YES	YES

Heteroskedasticity robust standard errors in parenthesis are adjusted for clustering at the school level
All regressions control for 10th grade test scores ethnicity, gender, LEP status and free or reduced lunch status.